Newton's problem of minimal resistance under the single-impact assumption

Alexander Plakhov*

Abstract

A parallel flow of non-interacting point particles is incident on a body at rest. When hitting the body's surface, the particles are reflected elastically. Assuming that each particle hits the body at most once (SIC condition), the force of resistance of the body along the flow direction can be written down in a simple analytical form.

The problem of minimal resistance within this model was first considered by Newton [15] in the class of bodies with a fixed length M along the flow direction and with a fixed maximum orthogonal cross section Ω, under the additional conditions that the body is convex and rotationally symmetric. Here we solve the problem (first stated in [5]) for the wider class of bodies satisfying SIC and with the additional conditions removed. The scheme of solution is inspired by Besicovitch's method of solving the Kakeya problem [1]. If Ω is a disc, the decrease of resistance as compared with the original Newton problem is more than twofold; the ratio tends to 2 as $M \to 0$ and to 20.25 as $M \to \infty$. We also prove that the infimum of resistance is 0 for a wider class of bodies with both single and double reflections allowed.

Mathematics subject classifications: 49Q10, 49K30

Key words and phrases: Newton's problem of minimal resistance, shape optimization, Kakeya problem, billiards.

*Center for R&D in Mathematics and Applications, Department of Mathematics, University of Aveiro, Portugal and Institute for Information Transmission Problems, Moscow, Russia

1 Introduction

1.1. Consider a bounded domain with piecewise smooth boundary (a *body*) B in Euclidean space $\mathbb{R}^3$ and a parallel flow of point particles with unit velocity incident on B. If a particle hits the body at a regular point of the boundary ∂B, it is reflected according to the elastic (billiard) law. A particle can make several reflections from the body. The particles do not interact with each other.

Under some additional assumptions (if, for example, the body B is convex) and knowing the flow density, it is possible to determine the force of pressure of the flow on the body. This force is usually called the *force of resistance*. One is traditionally interested in finding the body, in a prescribed class of bodies, that minimizes the projection of this force on the flow direction. This projection is also called the *resistance*.

Remarkably, this simple mechanical model is a source of various problems from different areas of mathematics. First stated by Newton [15] in a class of convex axisymmetric bodies, the problem of minimal resistance became one of the problems that gave origin to the calculus of variations. With the symmetry assumption removed, for various classes of convex bodies one comes to unusual and interesting multidimensional variational problems. They have been intensively studied in 1990s and 2000s (see [2]-[14]).

The condition of convexity guarantees absence of multiple collisions and allows one to write down the problem in a convenient analytical form. In the case of nonconvex bodies multiple reflections may occur, and one often needs to use methods of the billiard theory [16, 20, 21]. If, additionally, it is allowed to vary the direction v of the flow and one is interested in minimizing the resistance averaged over v, one comes to interesting problems related to optimal mass transfer [18, 17, 21].

Here we are going to study several problems of minimal resistance for bodies that are (generally) non-symmetric and nonconvex, but satisfy the so-called *single* (or *double*) *impact condition* (SIC or DIC): a particle of the flow cannot make more than one reflection (two reflections) from the body.

The single impact condition assures that the standard analytic formula for the resistance is preserved. Convex bodies obviously satisfy this condition, but not only they: if, for instance, a normal vector at each regular point of the part of ∂B faced to the flow makes an angle smaller than $\pi/6$ with the flow direction, then B satisfies SIC.

In brief, we consider several classes of bodies with a fixed length M along the direction of the flow and fixed maximum orthogonal cross section $\Omega \subset \mathbb{R}^2$. (It is assumed that Ω is convex.) We find the infimum of resistance $\phi(\Omega, M) > 0$ in the class of bodies satisfying SIC, and prove that the infimum is zero in the wider class of bodies satisfying DIC. Further, we consider bodies that are subgraphs of functions $u : \bar{\Omega} \to [0,\, M]$ and prove that the minimum resistance of such bodies coincides with $\phi(\Omega, M)$, independently of the restriction on the number of reflections.

Notice that the condition SIC and the related problem of minimal resistance were first stated in [5] and further discussed in [7]-[12].

1.2. Consider an orthonormal reference system $x_1,\, x_2,\, z$ in $\mathbb{R}^3$ and denote $x = (x_1, x_2)$. We assume that the flow falls vertically downward with velocity $v = (0, 0, -1)$.

Definition 1. Let Ω be a convex open bounded set in $\mathbb{R}^2$ and $u : \bar{\Omega} \to \mathbb{R}$ a piecewise smooth function. We say that u satisfies the *single impact condition* (SIC), if for any regular point $x \in \Omega$ and any $t > 0$ such that $x - t\nabla u(x) \in \bar{\Omega}$,

$$\frac{u(x - t\nabla u(x)) - u(x)}{t} \le \frac{1}{2}(1 - |\nabla u(x)|^2). \tag{1}$$

Remark 1. To each function u one naturally assigns the 3-dimensional body B_u bounded above by the graph of u and below by the horizontal plane $z = 0$,

$$B_u = \{(x, z) :\, x \in \bar{\Omega},\ 0 \le z \le u(x)\} \subset \mathbb{R}^3. \tag{2}$$

The single impact condition means that the trajectory of each particle after a reflection from the graph of u is situated above the graph. Equivalently, particles of the flow make no more than one reflection from the body B_u.

One easily calculates that the momentum imparted to the body by a particle hitting the body at a point $(x, u(x))$ equals $2\mu(1+|\nabla u(x)|^2)^{-1}(\nabla u(x),\, -1)$, where μ is the mass of the particle. Assume that u satisfies SIC (and therefore each particle hits the body only once). Summing up the momenta imparted by all incident particles per unit time, one concludes that the vertical component of the body's resistance force equals $-2\rho F(u)$, where ρ is the flow density and

$$F(u) = \int_\Omega \frac{dx}{1 + |\nabla u(x)|^2}. \tag{3}$$

The value $F(u)$ (3) is called the *resistance* of u. Note that if SIC is not satisfied, formula (3) has no physical meaning.

Definition 2. Let $M > 0$. We denote by $S_{\Omega,M}$ the class of functions $u : \bar{\Omega} \to \mathbb{R}$ satisfying SIC and such that $0 \le u(x) \le M$ for all x.

The following problem naturally arises.

Problem 1. *Find* $\inf_{u \in S_{\Omega,M}} F(u)$.

Remark 2. Problem 1 was stated in 1995 in [5] and has remained open since then. Here we provide the solution.

Remark 3. It is possible to define the condition SIC and state (and solve) the corresponding minimization problem for *nonconvex* domains Ω. However, this consideration would lead to technical complications in definitions and proofs, and therefore we restrict ourselves to convex domains.

Take the body B_u (2) corresponding to a piecewise smooth (not necessarily satisfying SIC) function $u : \bar{\Omega} \to \mathbb{R}$ and consider the billiard in $\mathbb{R}^3 \setminus B_u$. Let a billiard particle initially move according to $x(t) = x,\ z(t) = -t$, then make (maybe none) several reflections at regular points of ∂B_u, and finally move freely; we denote its final velocity by $v^+(x; u) = (v_1^+(x; u),\ v_2^+(x; u),\ v_3^+(x; u))$. If $x \not\in \bar{\Omega}$, one obviously has $v^+(x; u) = v$ (recall that $v = (0, 0, -1)$).

Definition 3. We say that the billiard scattering is *regular*, if the function $v^+(\cdot\,; u)$ is defined on a full-measure subset of $\mathbb{R}^2$ and is measurable. We denote by $U_{\Omega,M}$ the class of piecewise smooth functions $u : \bar{\Omega} \to \mathbb{R}$ such that
(a) $0 \le u(x) \le M$ for all $x \in \bar{\Omega}$ and
(b) the corresponding billiard scattering is regular.

Remark 4. It is easy to provide a function u that does not satisfy (b). Suppose that a part of graph(u) is a piece of a paraboloid of rotation whose focus coincides with a singular point of graph(u). Then the function $v^+(\cdot\,; u)$ is not defined in the projection of that piece of paraboloid on the x-plane.

The resistance of $u \in U_{\Omega,M}$ is defined by

$$F(u) = \int_\Omega \frac{1 + v_3^+(x; u)}{2}\, dx. \tag{4}$$

This formula also has a physical meaning. Indeed, a particle with mass μ that initially moves according to $x(t) = x,\ z(t) = -t$ and then makes several reflections from the body B_u, as a result imparts the momentum $\mu v - \mu v^+(x; u)$ to the body, and the projection of this momentum on the direction of v is

$$\mu\langle v - v^+(x; u),\ v\rangle = \mu(1 + v_3^+(x; u));$$

here and in what follows $\langle \cdot\,, \cdot \rangle$ means the scalar product. Summing all the imparted momenta, one finds that the vertical component of the body's resistance force to a flow with constant density ρ equals $-2\rho F(u)$.

One has $S_{\Omega,M} \subset U_{\Omega,M}$, and formulae (3) and (4) give the same value for $u \in S_{\Omega,M}$. Indeed, if $u \in S_{\Omega,M}$ then $v^+(x;u)$ is defined for all regular points x of u and, moreover, can be determined explicitly. Namely, a normal vector to graph(u) at $(x, u(x))$ is $n = (\nabla u(x), -1)$, and

$$v^+(x;u) = v - \frac{2\langle v,\, n\rangle}{|n|^2}\, n = \frac{1}{1 + |\nabla u(x)|^2}\, \big(-2\nabla u(x),\ 1 - |\nabla u(x)|^2 \big).$$

Therefore $(1 + v_3^+(x;u))/2 = 1/(1 + |\nabla u(x)|^2)$, and so, the integrals in the right hand sides of (3) and (4) coincide. This means that the value $F(u)$ is well defined.

We have the following problem.

Problem 2. *Find* $\inf_{u \in U_{\Omega,M}} F(u)$.

From the mechanical point of view, it makes sense to consider a wider class of bodies whose surface faced to the flow is not necessarily the graph of a function. For a 3-dimensional body B we again consider the billiard in $\mathbb{R}^3 \setminus B$ and analogously define the notion of regular scattering in terms of the final velocity $v^+(x;B) = (v_1^+(x;B),\, v_2^+(x;B),\, v_3^+(x;B)),\ x \in \mathbb{R}^2$.

Definition 4. Denote by $B_{\Omega,M}$ the class of bounded sets B in $\mathbb{R}^3$ with piecewise smooth boundary (bodies) such that

(a) $\bar\Omega \times \{0\} \subset B \subset \bar\Omega \times [0,\, M]$;

(b) the corresponding billiard scattering is regular.

The resistance of a body $B \in B_{\Omega,M}$ is defined by

$$R(B) = \int_\Omega \frac{1 + v_3^+(x;B)}{2}\, dx. \tag{5}$$

Problem 3. *Find* $\inf_{B \in B_{\Omega,M}} R(B)$.

Definition 5. For a body $B \in B_{\Omega,M}$ we define the function $u_B : \bar\Omega \to \mathbb{R}$ by

$$u_B(x) = \sup\{z : (x,z) \in B\}$$

(see Fig. 1).

The points of the first collision of the flow particles with the body lie on graph(u_B). If u_B satisfies SIC then the trajectory of the particle after the first collision lies above graph(u_B), and therefore has no further collisions with B.

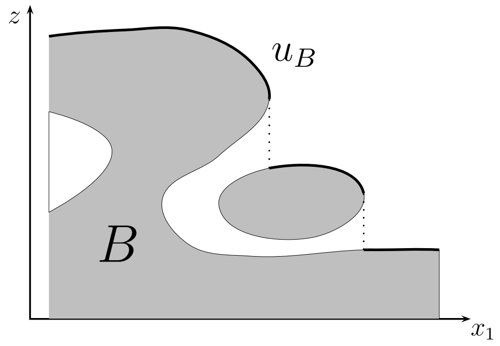

Figure 1: The cross section of a body B and of graph(u_B) by a vertical plane $x_2 = $const.

Remark 5. The class of bodies $B_{\Omega,M}$ is, in a sense, "larger" than the class of functions $U_{\Omega,M}$. In particular, the composition of mappings $u \mapsto B_u \mapsto u_{B_u}$ is the identity of $U_{\Omega,M}$, however the composition $B \mapsto u_B \mapsto B_{u_B}$ is not the identity of $B_{\Omega,M}$, but rather a projection onto a proper subset of $B_{\Omega,M}$.

Definition 6. We denote by $\Sigma_{\Omega,M}$ the class of bodies $B \in B_{\Omega,M}$ such that u_B satisfies SIC.

The following proposition states that the infima of resistance over *bodies* satisfying SIC and over *functions* satisfying SIC coincide.

Proposition 1. $\inf_{B\in\Sigma_{\Omega,M}} R(B) = \inf_{u\in S_{\Omega,M}} F(u)$.

Proof. It suffices to note that

(i) if $B \in \Sigma_{\Omega,M}$ then $u_B \in S_{\Omega,M}$ and $R(B) = F(u_B)$, and therefore $\inf_{B\in\Sigma_{\Omega,M}} R(B) \le \inf_{u\in S_{\Omega,M}} F(u)$;

(ii) if $u \in S_{\Omega,M}$ then $B_u \in \Sigma_{\Omega,M}$ and $F(u) = R(B_u)$, and therefore $\inf_{u\in S_{\Omega,M}} F(u) \le \inf_{B\in\Sigma_{\Omega,M}} R(B)$. $\square$

Actually, Problem 3 was solved in the book [21] (Section 2.1.2) in the case where Ω is a disc. The method used there can be easily generalized to arbitrary Ω to show that $\inf_{B\in B_{\Omega,M}} R(B) = 0$. In other words, if one allows multiple reflections, the resistance can be made arbitrarily small. It is then natural to fix the maximal allowed number of reflections and study the corresponding problem. Surprisingly enough, if single and double reflections are only allowed, the infimum of resistance equals zero.

Definition 7. We say that a body $B \in B_{\Omega,M}$ satisfies the *double impact condition* (DIC), if each incident particle with the initial velocity $v = (0, 0, -1)$ has no more than two reflections from B. The class of connected bodies satisfying DIC is denoted by $D_{\Omega,M}$.

Problem 4. *Find* $\inf_{B \in D_{\Omega,M}} R(B)$.

It is not difficult to find a positive lower bound for the resistance of a function $u \in S_{\Omega,M}$. To that end, put $d(x) = \text{dist}(x, \partial\Omega)$ and define

$$\phi(\Omega, M) = \int_\Omega \frac{1}{2}\left(1 - \frac{M}{\sqrt{M^2 + d^2(x)}}\right) dx. \tag{6}$$

It was first noticed in [5] that

$$\inf_{u \in S_{\Omega,M}} F(u) \geq \phi(\Omega, M). \tag{7}$$

For the sake of the reader's convenience, here we reproduce the proof of (7).

Let a particle have a (single) reflection at a regular point (x, z) of graph(u). It may further happen that the particle does or does not intersect the horizontal plane $z = 0$. In the former case let $(x', 0)$ be the point of intersection; then the final velocity is

$$v^+(x; u) = (x' - x,\ -z)/\sqrt{z^2 + |x' - x|^2}.$$

We have $x' \notin \Omega$ and therefore $|x - x'| \geq d(x)$, and $0 \leq z \leq M$; therefore

$$v_3^+(x; u) \geq -M/\sqrt{M^2 + d^2(x)}.$$

In the latter case we have $v_3^+(x; u) \geq 0$. In both cases the integrand in (4) is greater than or equal to $\frac{1}{2}\,(1 - M/\sqrt{M^2 + d^2(x)})$, and so, $F(u) \geq \phi(\Omega, M)$.

The following theorems provide solutions for problems 1 – 4.

Theorem 1.

$$\inf_{u \in S_{\Omega,M}} F(u) = \phi(\Omega, M). \tag{8}$$

Theorem 2.

$$\inf_{u \in U_{\Omega,M}} F(u) = \phi(\Omega, M) = \inf_{B \in \Sigma_{\Omega,M}} R(B). \tag{9}$$

Theorem 3.

$$\inf_{B \in B_{\Omega,M}} R(B) = 0 = \inf_{B \in D_{\Omega,M}} R(B). \tag{10}$$

These results are counterintuitive. Indeed, the statement of Theorem 1 implies that the graph of a nearly optimal function looks like a plateau with height M. The surface of the plateau is complicated and greatly inclined, with the angle of inclination being typically greater than 45^0. The plateau is crossed by a huge number of narrow deep valleys, but the total area covered by the valleys is small. The reflected particles further move along the valleys, and their density in the valleys is very high.

Theorem 1 provides the answer in the case where (a) *the front part of the body surface is the graph of a function* and (b) *the single impact condition is imposed*. If one of the conditions, (a) or (b), is removed (and thus a larger class of bodies is considered), the infimum remains the same, as indicated in Theorem 2. However, if both conditions are removed, the infimum is zero. It remains zero even if only single and double reflections are allowed. This is the claim of Theorem 3.

Note that a similar problem concerning minimization of specific resistance in hollows was considered in [22].

1.3. It is interesting to compare the minimizers in the four main classes of functions studied so far. We take the unit disc $\Omega = \Omega_0$ and consider the functions $u : \bar{\Omega}_0 \to \mathbb{R}$, $0 \le u \le M$ satisfying SIC, with the following additional conditions imposed:

($\mathrm{P_{SC}}$) u is radially symmetric and concave (the case considered by Newton [15]);

($\mathrm{P_C}$) u is concave [2, 5, 6, 13];

($\mathrm{P_S}$) u is radially symmetric [6, 7];

(P) no additional conditions on u (the present paper).

The minimizer exists in the classes ($\mathrm{P_{SC}}$), ($\mathrm{P_C}$), ($\mathrm{P_S}$) and does not exist in the class (P).

The optimal shapes in the classes ($\mathrm{P_{SC}}$), ($\mathrm{P_C}$), and ($\mathrm{P_S}$) with $M = 1$ are depicted in Figs. 2 (a)–(c). Figure 2 (c) is borrowed from [23]. Several trajectories of flow particles are also shown there.

Nearly optimal shapes in the class (P) are extremely complicated and not easy to depict. In Fig. 2 (d) a very schematic representation of a central vertical cross section of such a shape is given, also with $M = 1$. The particles shown in the figure after the reflection leave the plane of cross section and then move along narrow valleys (which are not shown); therefore their trajectories after the reflection are shown dashed.

The value of a nearly optimal function $u(x)$ in Problem (P) is typically close to the maximum value M, and $\nabla u(x)$ is typically close to

$$-\left(\frac{M}{1-|x|} + \sqrt{1 + \frac{M^2}{(1-|x|)^2}}\right)\frac{x}{|x|}.$$

On a subset of Ω_0 with small area we have $u(x) = 0$ and $\nabla u(x) = 0$. The infimum of resistance is given by formula (6), which in this case takes the form

$$\phi(\Omega_0, M) = \pi \int_0^1 \Big(1 - \frac{M}{\sqrt{M^2 + (1-r)^2}} \Big)\, r\, dr. \tag{11}$$

In the following table the values of minimal resistance are provided for Problems ($\mathrm{P_{SC}}$), ($\mathrm{P_C}$), and (P) with the values $M = 0.4, 0.7, 1, 1.5$. The data for the first two problems are taken from [13], and for the last one are calculated by formula (11).

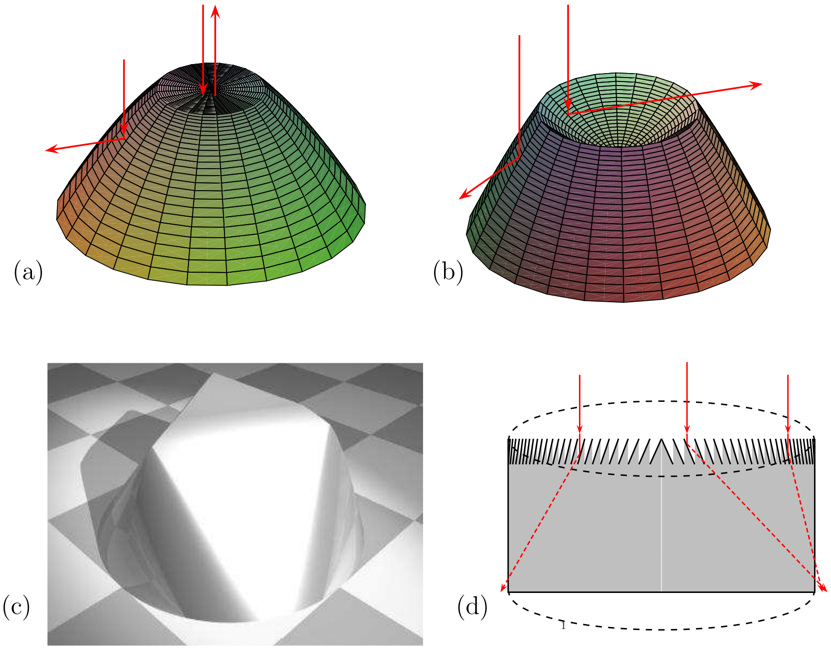

Figure 2: Optimal shapes for Problems ($\mathrm{P_{SC}}$), ($\mathrm{P_C}$), and ($\mathrm{P_S}$) are shown in Figs. (a)–(c). A schematic representation of a central vertical cross section of a nearly optimal body for Problem (P) is given in Fig. (d).

M	$\mathrm{P_{SC}}$	$\mathrm{P_C}$	P
1.5	0.75	0.70	0.05
1	1.18	1.14	0.10
0.7	1.57	1.55	0.18
0.4	2.11	2.11	0.35

Note that the values for Problem (P_C) were calculates with more precision in the recent paper [23]. I could not find in the literature numerical values of minimal resistance concerning Problem (P_S).

As $M \to 0$, the infimum of resistance goes to π in Problems (P_SC) and (P_C), and to $\pi/2$ in Problem (P). That is, the gain in our case is twofold.

As $M \to \infty$, the infimum in Problem (P_SC) is $\frac{27}{32}\pi M^{-2}(1 + o(1))$, and in (P) is $\frac{1}{24}\pi M^{-2}(1+o(1))$, that is, the gain in our case is more than twentyfold, as compared with Newton's case.

This improvement seems fantastic, but it is achieved at the expense of huge complication of optimal shapes. Our method does not allow to design applicable shapes, and in this sense it merely provides a result of existence. Even to get shapes with the resistance equal to or smaller than the minimal resistance in Newton's case, one needs to use details of the construction much smaller than the size of atoms.

The rest of the paper is organized as follows. The proof of Theorem 1 is given in Section 2, and the proofs of Theorems 2 and 3 are given in Section 3.

2 Proof of Theorem 1

2.1 The basic construction

Take a trapezoid $AMNB$ and assume that $|MN| < |AB|$. Let O be the point of intersection of the lines AM and BN (see Fig. 3).

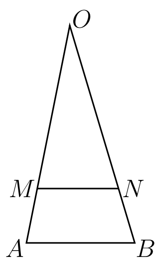

Figure 3: An elementary pair.

Let $d = \max\{|OA|, |OB|\}$, $d_0 = \text{dist}(O, \square AMNB)$, $h > 0$, and let p be the (unique) positive value satisfying

$$h = \frac{d^2 - p^2}{2p}.$$

It is easy to check that

$$p = \sqrt{d^2 + h^2} - h.$$

For any $x \in \mathbb{R}^2$ denote by $r(x)$ the distance between x and O, and let $\triangle$ be the open triangle MON, and $\square$ be the open trapezoid $\square AMNB$.

Definition 8. The trapezoid $\square$ is called an *elementary mirror* and the triangle $\triangle$, the corresponding *elementary valley*. The pair $(\triangle, \square)$ is called an *elementary pair*, and the point O, the *focus of this pair*.

The *elementary h-function* $u = u_{h,\square} : \overline{\triangle AOB} \to \mathbb{R}$ corresponding to the elementary pair is defined by

$$u(x) = \begin{cases} \frac{r^2(x) - p^2}{2p}, & \text{if } x \in \square; \\ 0 & \text{otherwise.} \end{cases}$$

That is, u equals 0 in the triangle $\triangle$ and on the boundary of the triangle and the trapezoid $\square$.

The *ratio* $\varkappa = \varkappa(\square)$ of the elementary pair is defined by

$$\varkappa = \frac{d - d_0}{d}.$$

Lemma 1. *For $x \in \square$ one has*

$$h - \varkappa\,(\sqrt{d^2 + h^2} + h) < u(x) < h, \tag{12}$$

$$\frac{1}{2}\left(1 - \frac{h}{\sqrt{d^2 + h^2}}\right) < \frac{1}{1 + |\nabla u(x)|^2} < \frac{(1 - \varkappa)^{-2}}{2}\left(1 - \frac{h}{\sqrt{d^2 + h^2}}\right). \tag{13}$$

Proof. Since $d_0 < r(x) < d$, we have

$$\frac{d_0^2 - p^2}{2p} < u(x) < \frac{d^2 - p^2}{2p}. \tag{14}$$

The right hand side of (14) equals h, and the left hand side equals

$$\frac{d^2 - p^2}{2p} - \frac{d^2 - d_0^2}{2p} = h - \frac{d + d_0}{2} \cdot \frac{d - d_0}{\sqrt{d^2 + h^2} - h}$$

$$> h - d \cdot \frac{d\varkappa}{\sqrt{d^2 + h^2} - h} = h - \varkappa(\sqrt{d^2 + h^2} + h).$$

We have $|\nabla u(x)| = r(x)/p < d/p$, therefore

$$\frac{1}{1 + |\nabla u(x)|^2} > \frac{p^2}{p^2 + d^2} = \frac{1}{2}\left(1 - \frac{h}{\sqrt{d^2 + h^2}}\right).$$

On the other hand, $|\nabla u(x)| > d_0/p = (1-\varkappa)d/p$, therefore

$$\frac{1}{1+|\nabla u(x)|^2} < \frac{p^2}{p^2+(1-\varkappa)^2 d^2} < \frac{(1-\varkappa)^{-2}p^2}{p^2+d^2} = \frac{(1-\varkappa)^{-2}}{2}\left(1-\frac{h}{\sqrt{d^2+h^2}}\right).$$

Lemma 1 is proved. $\square$

By the first inequality in (12), the function $u_{h,\square}$ is non-negative, if

$$\varkappa(\square) \le \frac{1}{1+\sqrt{1+d^2/h^2}}. \tag{15}$$

Lemma 2. *A nonnegative elementary h-function $u = u_{h,\square}$ satisfies SIC.*

Proof. Let us give both geometrical and analytical proofs of the lemma. Geometrically, a particle that initially projects on the trapezoid, is reflected from the graph of u and then goes downward along a line through the point $(O, 0)$. The section of the graph of u by the vertical plane containing the trajectory of the particle is shown bold in Fig. 4. If, on the other hand, the

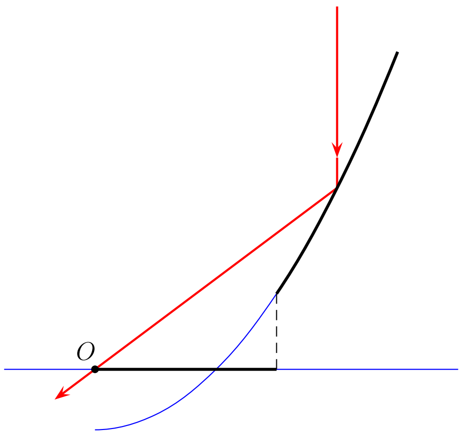

Figure 4: A vertical section of graph(u) (shown bold) and the trajectory of a particle.

particle initially projects on the triangle MON, it is reflected vertically.

The analytical proof is a little bit more involved. Put the origin at the point O; then for $x \in \square$ one has $u(x) = \frac{1}{2p}(|x|^2 - p^2)$ and $\nabla u(x) = x/p$, and condition SIC (1) reads as follows: for any $0 < t \le p$,

$$u\Big(\frac{p-t}{p}\,x\Big) - u(x) \le \frac{t}{2}\Big(1 - \frac{|x|^2}{p^2}\Big). \tag{16}$$

One has $\frac{p-t}{p}\,x \in \bar{\triangle} \cup \square$. If $\frac{p-t}{p}\,x \in \bar{\triangle}$, inequality (16) takes the form $-u(x) \le -\frac{t}{p}\,u(x)$, which is obviously true since $t \le p$ and $u(x) > 0$. If $\frac{p-t}{p}\,x \in \square$, inequality (16) takes the form

$$\frac{\left(\frac{p-t}{p}\right)^2 |x|^2 - |x|^2}{2p} \le \frac{t}{2}\left(1 - \frac{|x|^2}{p^2}\right),$$

which after some algebra reduces to the obvious inequality $(t-p)\,|x|^2/p^3 \le 1$.

If $x \in \triangle$, one has $u(x) = 0,\ \nabla u(x) = 0$, and $u\big(\frac{p-t}{p}\,x\big) = 0$ for all $0 < t \le p$, and condition SIC takes the form $0 \le 1/2$. Lemma 2 is proved. $\square$

The resistance of a nonnegative function $u_{h,\square}$ equals

$$F(u_{h,\square}) = |\triangle| + \int_\square \frac{dx}{1 + |\nabla u(x)|^2}; \tag{17}$$

here and below, $|\triangle|$ and $|\square|$ mean the areas of the triangle MON and the trapezoid $AMNB$, respectively. By the second inequality in (13), the integrand in the right hand side of (17) does not exceed $\frac{1}{2}(1-\varkappa)^{-2}\left(1 - h/\sqrt{d^2+h^2}\right)$, so the resistance of u is estimated as follows:

$$F(u_{h,\square}) \le |\triangle| + \frac{(1-\varkappa)^{-2}}{2}\left(1 - \frac{h}{\sqrt{d^2+h^2}}\right) \cdot |\square|.$$

Lemma 3. *Consider a finite collection of elementary pairs $\triangle_i$, $\square_i$ with poles at O_i. Choose positive values h_i so that the elementary h_i-functions $u_i = u_{h_i,\square_i} : \overline{\triangle_i \cup \square_i} \to \mathbb{R}$ are nonnegative. Let a convex domain Ω be such that $\Omega \subset \cup_i \overline{(\triangle_i \cup \square_i)}$ and all O_i lie outside Ω. Let the function $u : \bar{\Omega} \to \mathbb{R}$ be defined by $u(x) = \min_i u_i(x)$, where the infimum is taken over those i for which $x \in \overline{\triangle_i \cup \square_i}$. Then u satisfies SIC.*

The proof of Lemma 3 is a simple consequence of the definitions and is left to the reader.

The following important lemma will be proved in the next two subsections.

Lemma 4. *For any $\varepsilon > 0$ there exist a finite family of elementary pairs $\triangle_i$, $\square_i$ with ratios $\varkappa_i$ and with foci at O_i such that*

(A) $\Omega \subset \cup_i \overline{(\triangle_i \cup \square_i)}$;

(B) $|\cup_i \triangle_i| < \varepsilon$;

(C) $\varkappa_i < \varepsilon$;

(D) for each i, $O_i \notin \Omega$;

(E) for each i and $x \in \square_i$, $|x - O_i| < dist(x, \partial\Omega) + \varepsilon$.

The set

$$V = (\cup_i \triangle_i) \cap \Omega$$

is called *the valley* of the family, and its complement $\Omega \setminus V$, *the mirror* of the family.

Notice that the family of elementary pairs $\triangle_i = \triangle_i(\varepsilon)$, $\square_i = \square_i(\varepsilon)$ indicated in this lemma depends on ε.

Let us now derive Theorem 1 from Lemma 4. Consider the elementary M-functions $u_{i,\varepsilon} = u_{M,\square_i(\varepsilon)} : \overline{\triangle_i \cup \square_i} \to \mathbb{R}$ and define the function $u_\varepsilon : \bar{\Omega} \to \mathbb{R}$ by $u_\varepsilon(x) = \inf_i u_{i,\varepsilon}(x)$, where the infimum is taken over those i for which $x \in \overline{\triangle_i \cup \square_i}$.

Lemma 5. *For ε sufficiently small we have $u_\varepsilon \in S_{\Omega,M}$ and $F(u_\varepsilon) < \phi(\Omega, M) + O(\varepsilon)$, $\varepsilon \to 0$. Further, there exists a finite set of points $O_i = O_i^\varepsilon \notin \Omega$ in the ε-neighborhood of Ω and a domain $V \subset \Omega$ with area $|V| < \varepsilon$ such that each incident particle corresponding to $x \in V$ is reflected vertically, and each particle corresponding to a regular point $x \in \Omega \setminus V$ after the reflection passes through one of the points $(O_i, 0)$ on the x-plane.*

Proof. Property (E) implies that $d_i \le \mathrm{diam}(\Omega) + \varepsilon$. Taking into account property (C), we conclude that for ε sufficiently small and arbitrary i,

$$\varkappa_i < \varepsilon < \frac{1}{1 + \sqrt{1 + d_i^2/M^2}},$$

and so, inequality (15) is satisfied for the function $u_{i,\varepsilon}$. This means that $0 \le u_{i,\varepsilon} \le M$, and by Lemma 2, $u_{i,\varepsilon}$ satisfies SIC. By properties (A) and (D) and Lemma 3, u_ε also satisfies SIC. Therefore $u_\varepsilon \in S_{\Omega,M}$.

If $x \in V$, the particle is reflected vertically. If x is a regular point of $\Omega \setminus V$ then for some i we have the equality $u_\varepsilon = u_{i,\varepsilon}$ in a neighborhood of x, and so, $x \in \square_i$. The corresponding particle after the reflection passes through the point $(O_i, 0)$ on the x-plane. Using (D) and (E) and taking into account that $x \in \square \cap \Omega$, one easily concludes that $\mathrm{dist}(O_i, \Omega) < \varepsilon$. Indeed, the segment $[x, O_i]$ contains a point $y \in \partial\Omega$, and since $|x - y| \ge \mathrm{dist}(x, \partial\Omega)$ and $|x - O_i| <\mathrm{dist}(x, \partial\Omega) + \varepsilon$, we have

$$\mathrm{dist}(O_i, \Omega) \le |O_i - y| = |x - O_i| - |x - y| < \varepsilon.$$

Further, by Lemma 1

$$\frac{1}{1 + |\nabla u_\varepsilon(x)|^2} = \frac{1}{1 + |\nabla u_{i,\varepsilon}(x)|^2} < \frac{(1 - \varkappa_i)^{-2}}{2} \left(1 - \frac{M}{\sqrt{M^2 + d_i^2}} \right)$$

$$< \frac{(1-\varepsilon)^{-2}}{2}\left(1-\frac{M}{\sqrt{M^2+(d(x)+\varepsilon)^2}}\right)$$

(recall that $d(x) = \mathrm{dist}(x, \partial\Omega)$). Thus,

$$F(u_\varepsilon) = \int_\Omega \frac{dx}{1+|\nabla u_\varepsilon(x)|^2} < |V| + \frac{(1-\varepsilon)^{-2}}{2}\int_{\Omega\setminus V}\left(1-\frac{M}{\sqrt{M^2+(d(x)+\varepsilon)^2}}\right)dx.$$

By property (B), $|V| < \varepsilon$, and recalling equation (6) we conclude that $F(u) < \phi(\Omega, M) + O(\varepsilon)$. Lemma 5 is proved. □

The claim of Theorem 1 is a consequence of Lemma 5 and equation (7).

2.2 Proof of Lemma 4

Assume that we are given a bounded convex open set Ω and a positive value ε.

Consider a square lattice $\delta\mathbb{Z} \times \delta\mathbb{Z}$, the size δ of the lattice to be specified below, and take the (closed) squares $Q_1, \dots, Q_N$ of the lattice that have nonempty intersection with Ω. Consider also the lattice $2\delta\mathbb{Z} \times 2\delta\mathbb{Z}$ with double size and take the squares $\tilde{Q}_1$, $\tilde{Q}_2, \dots$ of this lattice that do not intersect Ω, $\tilde{Q}_i \cap \Omega = \emptyset$. For each square Q_i find the square $\tilde{Q}_i$ such that the distance between their centers is minimal (and change the numeration of $\tilde{Q}_i$ if necessary); see Fig. 5.

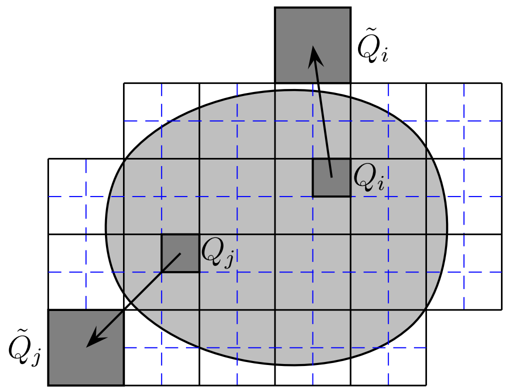

Figure 5: Covering of Ω by squares of the lattice $\delta\mathbb{Z} \times \delta\mathbb{Z}$ and two cases of correspondence between the squares Q_i and $\tilde{Q}_i$.

The correspondence $Q_i \mapsto \tilde{Q}_i$ is not necessarily injective; that is, it may happen that $\tilde{Q}_i = \tilde{Q}_j$ for $i \neq j$.

Choose δ small enough so that for each i and for any two points $x \in Q_i$, $\tilde{x} \in \tilde{Q}_i$,

$$|x - \tilde{x}| < \operatorname{dist}(x, \partial\Omega) + \varepsilon/2. \tag{18}$$

It suffices to take, for example, $\delta = \varepsilon/10$.

The following lemma will be proved in the next subsection. We use the notation $B_\omega(C)$ for the ball with radius ω centered at the point C.

Lemma 6. *For any triangle ABC there exists a finite family of elementary pairs $\triangle_i^\omega$, $\square_i^\omega$ depending on the parameter $\omega > 0$ with ratios $\varkappa_i^\omega$ and with foci at O_i^ω (the number of pairs in the family may depend on ω) such that*

(a) $B_\omega(C) \subset \overline{\cup_i (\triangle_i^\omega \cup \square_i^\omega)}$;

(b) $|\cup_i \triangle_i^\omega|/\omega^2 \to 0$ *as* $\omega \to 0$;

(c) $\max_i \varkappa_i^\omega \to 0$ *as* $\omega \to 0$;

(d) the foci O_i^ω belongs to the $\alpha(\omega)$-neighborhood of the segment AB, and

(e) $\cup_i \square_i^\omega \subset B_{\alpha(\omega)}(C)$, *where* $\alpha(\omega) \to 0$ *as* $\omega \to 0$.

Let us prove Lemma 4 using this lemma.

Fix i and define a triangle ABC so that the translates of Q_i by the vectors $\overrightarrow{CA}$ and $\overrightarrow{CB}$ lie in the interior of $\tilde{Q}_i$. This is possible, since the size of Q_i is smaller than that of $\tilde{Q}_i$.

Take an $n \in \mathbb{N}$ and divide Q_i into n^2 small squares; let $C_1, \ldots, C_{n^2}$ be their centers (see Fig. 6 (b)). The size of each square is

$$\omega = \operatorname{size}(Q_i)/n. \tag{19}$$

Take the family of elementary pairs $\{\triangle_i^\omega, \square_i^\omega\}$ defined by Lemma 6 for the triangle ABC, and for each $k = 1, \ldots, n^2$ consider the translate of this family by the vector $\overrightarrow{CC_k}$ (see Fig. 6 (b)). The corresponding translate of the triangle ABC will be denoted by $A_kB_kC_k$, and the union of translates of the family $\{\triangle_i^\omega, \square_i^\omega\}$ will be referred to as the *big family corresponding to Q_i*. We have the following.

(a) The union of elementary sets in the kth family contains the kth small square; therefore the union of elementary sets in the big family contains Q_i.

(b) By (19), the area of the union of elementary triangles in the big family is not greater than

$$n^2 |\cup_i \triangle_i^\omega| = (\operatorname{side}(Q_i))^2 \frac{1}{\omega^2} |\cup_i \triangle_i^\omega| \to 0 \quad \text{as } \omega \to 0.$$

For n sufficiently large it is smaller than ε/N.

(c) For n sufficiently large (and therefore, ω sufficiently small), $\varkappa_i^\omega < \varepsilon$.

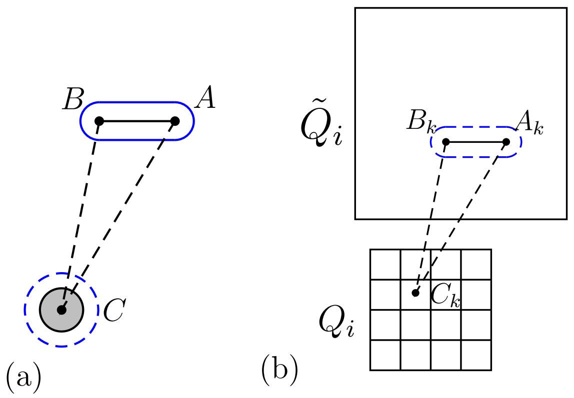

Figure 6: (a) The union of elementary sets in the family contains the shaded circle, and all the trapezoids are contained in the circle bounded by a dashed line. The focal set of the family is contained in the neighborhood of AB shown by a dashed line. (b) A pair of squares Q_i, $\tilde{Q}_i$ and a translate of $\triangle ABC$ corresponding to a small square in Q_i.

(d) For ω sufficiently small, not only the translates of Q_i by $\overrightarrow{CB}$ and $\overrightarrow{CA}$ belong to $\tilde{Q}_i$, but also their $\alpha(\omega)$-neighborhoods. This implies that the $\alpha(\omega)$-neighborhoods of the segments $\overrightarrow{A_kB_k}$, $k = 1, \ldots, n^2$ also belong to $\tilde{Q}_i$. Thus, for n sufficiently large (and correspondingly ω sufficiently small), all foci of the big family belong to $\tilde{Q}_i$, and therefore do not belong to Ω.

(e) For ω sufficiently small, $\alpha(\omega) < \varepsilon/4$, and therefore the union of the trapezoids of the big family belongs to the $(\varepsilon/4)$-neighborhood of Q_i. Take a point x in a trapezoid of the big family and let O be the corresponding focus. Then there exists a point $x' \in Q_i$ such that $|x - x'| < \varepsilon/4$ and therefore

$$\operatorname{dist}(x', \partial\Omega) < \operatorname{dist}(x, \partial\Omega) + \varepsilon/4.$$

Since (for ω sufficiently small) $O \in \tilde{Q}_i$, by (18) we have

$$|x' - O| < \operatorname{dist}(x', \partial\Omega) + \varepsilon/2,$$

and thus,

$$|x - O| < |x - x'| + |x' - O| < \varepsilon/4 + \operatorname{dist}(x', \partial\Omega) + \varepsilon/2 < \operatorname{dist}(x, \partial\Omega) + \varepsilon.$$

Take the union of the big families corresponding to all Q_i, $i = 1, \ldots, n^2$; it follows from (a)–(e) that it satisfies the conditions (A)–(E). Lemma 4 is proved.

2.3 Proof of Lemma 6

The family to be constructed is a two-level hierarchy. The construction of families of the 1st order is actually a slight modification of the construction used by Besicovitch [1] to solve the Kakeya needle problem. These families have been used by several authors for various purposes. Fefferman [11] used them to disprove the so-called disc multiplier conjecture in harmonic analysis. The families served as a part of the construction of simple connected and uniformly bounded Kakeya sets in the paper of Cunningham [10]. They were also used in [22] to solve the problem of minimal Newtonian resistance in a hollow. For the reader's convenience, here we describe the construction.

Let $0 < k < 1$. Take a triangle MON and extend the sides OM and ON beyond the point O to obtain the segments MM' and NN', with

$$|OM'| = k|OM| \quad \text{and} \quad |ON'| = k|ON|.$$

Joining the points M' and N' with the midpoint D of MN, we obtain two triangles $MM'D$ and $NN'D$ (see Fig. 7). The procedure of k-doubling consists in substituting the original triangle MON with the two triangles $MM'D$ and $NN'D$. If the height of $\triangle MON$ is h, then the heights of the new triangles are both equal to $(1 + k)h$.

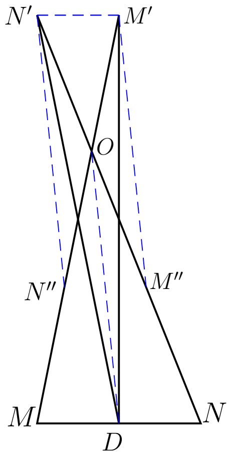

Figure 7: The procedure of doubling: the triangle MON is substituted with the two triangles $MM'D$ and $NN'D$.

It is easy to estimate the increase of the total area, $|\triangle MM'D \cup \triangle NN'D| - |\triangle MON|$, as a result of doubling. Draw two lines parallel to OD through M' an N', and denote by M'' and N'' the points of intersection of these lines with ON and OM, respectively. We have

$$|\triangle ON'N''| = |\triangle OM'M''| = k^2 |\triangle MON|,$$

therefore

$$|\triangle MM'D \cup \triangle NN'D| - |\triangle MON| < |\triangle ON'N'' \cup \triangle OM'M''| = 2k^2 |\triangle MON|.$$

Let us now apply the procedure of doubling successively several times, starting from the triangle MON. Assume that the height of MON is 1 and the length of the base is d. At the mth step, $m = 1, 2, \dots$ we apply k_m-doubling with $k_m = 1/m$. After m steps we get 2^m triangles with height $m+1$ and with the length of base $2^{-m}d$.

Let S_m be the area of the union of triangles after the mth step. One has $S_0 = |\triangle MON| = d/2$, and the total increase of the area at the mth step is smaller than $2^{m-1} \cdot 2\delta_m^2 \cdot 2^{-(m-1)} dm/2 = d/m$, therefore

$$S_m < S_{m-1} + d/m.$$

One easily concludes by induction that $S_m < d(\ln m + 3/2)$ for $m \geq 1$.

Now fix $l > 0$ and for $m = 1, 2, \dots$ define the family of 2^m elementary pairs $\triangle_i$, $\square_i$, where the triangles $\triangle_i$ coincide with the triangles obtained at the mth step of doubling, and all the trapezoids $\square_i$ have the same height l. Each trapezoid of the $(m-1)$th step generates two trapezoids of the mth step, which are disjoint and contained in the original one. This implies that for each m, the trapezoids of the mth step are mutually disjoint. The area of the union of trapezoids is greater than ld.

Definition 9. A family $\triangle_i$, $\square_i$ of 2^m elementary pairs at the mth step with $l = \sqrt{m}$ is called a *family of the 1st order*. The union $\cup_{i=1}^{2^m} \overline{(\triangle_i \cup \square_i)}$ is called a $(1,m)$-*set*. The (finite) set of foci of the elementary pairs is called the corresponding *focal set of the 1st order*. The original triangle MON is called the *generating triangle* for the family of the 1st order.

Several families of the 1st order with $m = 0, 1, 2, 3, 4$ are shown in Fig. 8.

Note in passing that all families of the 1st order with fixed m are linearly isomorphic. That is, for any two such families there exists a linear transformation that takes a set (the generating triangle, the ith triangle, the ith trapezoid, the ith focus, $i = 1, \dots, 2^m$) of the former family to the corresponding set of the latter one.

For the ratios of the elementary pairs in the family we have the estimate

$$\varkappa_i \leq \frac{\sqrt{m} + 2^{-m}d}{m + 1 + \sqrt{m}} \to 0 \quad \text{as } m \to \infty. \tag{20}$$

Further, for the areas of the union of triangles and the union of trapezoids we have

$$\frac{|\cup_{i=1}^{2^m} \triangle_i|}{|\cup_{i=1}^{2^m} \square_i|} \leq \frac{\ln m + 3/2}{\sqrt{m}} \to 0 \quad \text{as } m \to \infty. \tag{21}$$

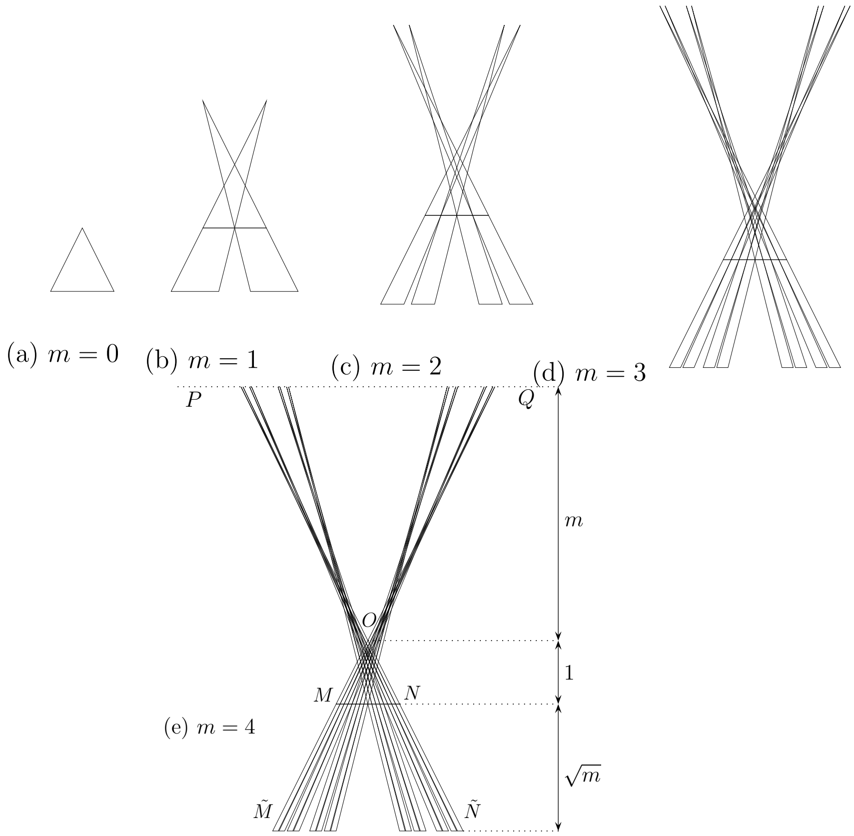

Figure 8: Families of the 1st order, with $m = 0, 1, 2, 3, 4$. In Fig. (e) the generating triangle of the family is MON, the set $\mathcal{C}$ is the triangle $\tilde{M}O\tilde{N}$, and the focal set lies on the line PQ.

Remark 6. The union of trapezoids $\cup_{i=1}^{2^m} \square_i$ is interpreted as the support of mirrors, and the union of triangles $\cup_{i=1}^{2^m} \triangle_i$ as the support of valleys in the family. Formula (21) means that as $m \to \infty$, the relative area of the support of valleys vanishes. The ratios $\varkappa_i$ govern the resistance of mirrors in the family. Roughly speaking, formula (20) indicates that as $m \to \infty$, the resistances of the mirrors can be made arbitrarily close to their smallest value.

Remark 7. The first idea that comes to mind is to prove Theorem 1 directly by putting a large number of small copies of $(1, m)$-sets inside Ω. That is, we first put inside Ω as large copy F_0 of a $(1, m)$-set as we can, then put several smaller copies $F_1, \ldots, F_p$ of $(1, m)$-sets inside the uncovered part $\Omega \setminus F_0$, then put even smaller copies inside $\Omega \setminus (F_0 \cup F_1 \cup \ldots \cup F_p)$, etc. The procedure is repeated until a sufficiently large part of Ω is covered. The focal sets of all the copies should lie outside Ω. The problem, however, is that the parameter m should go to infinity in this hierarchy of copies. As a result, the sets in the hierarchy become more and more complicated, and one cannot guarantee that the area of the uncovered part of Ω goes to zero.

We will use instead a more sophisticated construction. Take a $(1, m)$-set and consider the convex hall $\mathcal{C}$ of the union of the generating triangle MON and the trapezoids of the family. It is the triangle homothetic to the generating one, with the ratio $\sqrt{m}+1$ and with center of homothety at the vertex O of the generating triangle. The height of $\mathcal{C}$ equals $\sqrt{m}+1$. The focal set of the 1st order lies on the base of another triangle homothetic to $\triangle MON$, with the same center of homothety O and with the ratio m (see Fig. 8).

Notice that the set of trapezoids is very "sparse" in $\mathcal{C}$: the total area of the trapezoids is $\sqrt{m}d(1+o(1))$, $m \to \infty$, whereas the area of $\mathcal{C}$ is greater than $md/2$. Actually, the part of $\mathcal{C}$ occupied with the $(1, m)$-set is $\mathcal{C}$ minus the union of $2^m - 1$ angles with the vertices on the base of the generating triangle. These angles are called *angles associated with the* $(1, m)$*-set*; they do not mutually intersect.

Recall that each set obtained from a $(1, m)$-set by a linear transformation is again a $(1, m)$-set. Apply the following iterative procedure. At the first step take $2^m - 1$ sets obtained from the original $(1, m)$-set by linear transformations that take the generating triangle MON to triangles, with height 1 and with the base parallel to MN, inscribed in the $2^m - 1$ angles associated with the original $(1, m)$-set (so that the vertices of the resulting generating triangles coincide with the vertices of the associated angles). The union of all the 2^m sets (including the original one) contains the triangle homothetic to the original generating triangle MON with the ratio 2 and the same center O. In Fig. 9 there are shown the original $(1, 2)$-set and one of the three new $(1, 2)$-sets with the generating triangle inscribed into the central associated angle.

At the second step we repeat the procedure as applied to each of the new $(1, m)$-sets. As a result we obtain $(2^m - 1)^2$ copies of the original $(1, m)$-set that fit into the new $(2^m - 1)^2$ associated angles. The union of all the obtained sets (their total number is $1 + (2^m - 1) + (2^m - 1)^2$) contains the triangle homothetic to $\triangle MON$ with the ratio 3.

Note in passing that te height of the generating triangle of each copy obtained this way is 1, and the base is smaller than d. Therefore the ratios of all obtained elementary pairs satisfy (20).

We repeat this procedure $\lfloor \sqrt{m} \rfloor + 1$ times, where $\lfloor \ldots \rfloor$ means the inte-

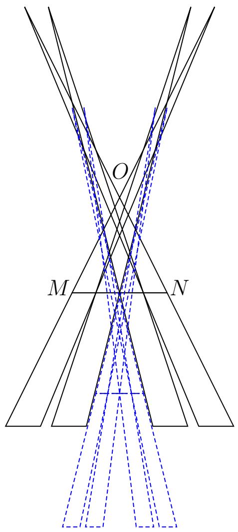

Figure 9: Starting the construction of a family of the 2nd order.

ger part. As a result, $\mathcal{C}$ will be contained in the union of all the obtained $(1,m)$-sets, including the original one and the ones obtained at the steps $1, 2, \ldots, \lfloor\sqrt{m}\rfloor + 1$.

Definition 10. The resulting family of elementary pairs is called a *family of the 2nd order*. The union of the obtained $(1,m)$-sets is called a $(2,m)$-*set*. The union of the corresponding focal sets is called the *focal set of the 2nd order*.

Note that the image of a family of the 2nd order under a linear transformation is again a family of the 2nd order.

Let us characterize in more detail the linear transformations that take the original $(1,m)$-set to the new ones at each step of the procedure. Each of these transformations can be decomposed into two ones. The first transformation preserves O and transforms the segment MN into itself. The second one is a translation that moves O into the triangle $\mathcal{C}$. In other words, it is the translation by a vector $\overrightarrow{OA}$, where $A \in \mathcal{C}$.

The focal set of the 2nd order F is the union of the original focal set of the 1st order and its images under these transformations. Let $[M'N']$ be the image of the segment $[MN]$ under the homothety with center at O and the ratio $-m$. The original focal set of the 1st order lies on $[M'N']$. Any other point of F is thus obtained from a point of the segment $[M'N']$ in the following way: first move it to another point of the segment and then translate it by a vector $\overrightarrow{OA}$, with $A \in \mathcal{C}$.

Thus, F can be characterized as follows. Put the origin at O and consider the algebraic sum $[M'N']+\mathcal{C}$ of the segment $[M'N']$ and the triangle $\mathcal{C}$. Then

$$F \subset [M'N'] + \mathcal{C}.$$

The domain $[M'N'] + \mathcal{C}$ is actually a trapezoid; it is shown bounded by a dashed line in Fig. 10. It belongs to an $O(\sqrt{m})$-neighborhood of the segment $[M'N']$.

By the construction, the $(2, m)$-set contains $\mathcal{C}$, and its elementary trapezoids are all contained in the triangle $2\mathcal{C}$ (where again the origin is at O); see Fig. 10.

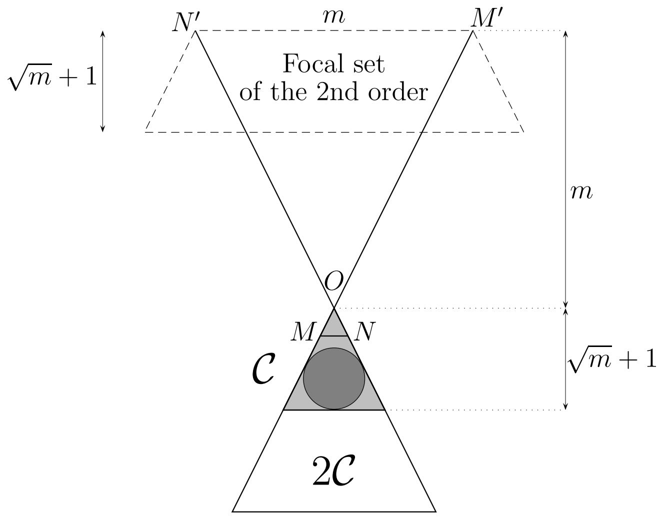

Figure 10: The $(2, m)$-set contains the shaded triangle $\mathcal{C}$. Its focal set is situated within the dashed line. The elementary trapezoids of the family are all contained in the triangle $2\mathcal{C}$.

The elementary trapezoids of the family do not mutually intersect and are contained in the triangle $2\mathcal{C}$ with height $2\sqrt{m} + 2$ and base $(2\sqrt{m} + 2)d$. Therefore the sum of their areas is smaller than $(2\sqrt{m} + 2)^2 d/2$. The family of elementary trapezoids is divided into several sub-families of the 1nd order. For each sub-family of trapezoids and the corresponding sub-family of elementary triangles, inequality (21) is valid. Summing over all sub-families, we conclude that the area of the union of elementary triangles in the family of the 2nd order is smaller than

$$\frac{\ln m + 3/2}{\sqrt{m}} \, \frac{(2\sqrt{m} + 2)^2 d}{2} = 2d\sqrt{m} \ln m(1 + o(1)), \quad m \to \infty. \tag{22}$$

The circle inscribed in $\mathcal{C}$ has the radius $r_1\sqrt{m}(1+o(1))$, and the minimal concentric circle containing $2\mathcal{C}$ has the radius $r_2\sqrt{m}(1+o(1))$, $m \to \infty$, where r_1 and r_2 are positive constants that depend only on the generating triangle.

Now introduce a positive parameter ω and let $m = \lfloor c^2/\omega^2 \rfloor$, the constant $c > 0$ to be specified below. Let the generating triangle MON be homothetic, with a negative ratio $-r$, to the triangle ABC indicated in Lemma 6. Then $\triangle M'ON'$ is homothetic to $\triangle ABC$ with the ratio mr. Apply to the family of the 2nd order the composition of two transformations. The first one is the translation that takes the center of the inscribed circle to O. (The translation distance is $O(\sqrt{m})$.) The second one is a homothety with the ratio $1/(mr)$ that takes O to C, M' to A, and N' to B. The image is another family of the 2nd order (let it be denoted by $\{\triangle_i^\omega, \square_i^\omega\}$) that satisfies the following properties.

(a) The union $\cup_i \overline{(\triangle_i^\omega \cup \square_i^\omega)}$ of the obtained elementary sets contains the circle centered at C with radius $\frac{\omega}{c}\frac{r_1}{r}(1+o(1))$, $\omega \to 0$. Take $c < r_1/r$; then for ω sufficiently small the union contains $B_\omega(C)$.

(b) Since the ratio of homothety is $1/(mr)$, the area $|\cup_i \triangle_i^\omega|$ of the union of triangles in the resulting family is smaller than $1/(mr)^2$ times the expression in (22). Taking into account that $1/m \sim \omega^2$, we conclude that

$$\frac{|\cup_i \triangle_i^\omega|}{\omega^2} \sim \omega \ln \frac{1}{\omega} \to 0 \quad \text{as } \omega \to 0.$$

(c) The ratio of an elementary pair is invariant under a homothety; therefore the ratios of all elementary pairs of the family satisfy (20). Thus, we have

$$\max_i \varkappa_i^\omega \le \frac{\sqrt{m} + 2^{-m}d}{m + 1 + \sqrt{m}} \sim \omega \to 0 \quad \text{as } \omega \to 0.$$

(d) The focal set of the family is contained in the image of the trapezoid $[M'N'] + \mathcal{C}$, which in turn belongs to an $O(1/\sqrt{m})$-neighborhood of the segment $[AB]$. Taking into account that $1/\sqrt{m} \sim \omega$, we conclude that the focal set belongs to the $O(\omega)$-neighborhood of $[AB]$.

(e) The union $\cup_i \square_i^\omega$ of the trapezoids of the family belongs to the image of $2\mathcal{C}$, which in turn belongs to the circle centered at C with radius $\frac{\omega}{c}\frac{r_2}{r}(1+o(1))$, $\omega \to 0$. That is, $\cup_i \square_i^\omega \subset B_{O(\omega)}(C)$.

Thus, Lemma 6 is proved.

3 Proofs of Theorems 2 and 3

3.1 Theorem 2

The second equality in (9) is a consequence of Theorem 1 and Proposition 1. It remains to prove the first one.

The trajectory of a billiard particle is naturally parameterized by the time t. Let a particle initially move according to $x(t) = x_0$, $z(t) = -t$ ($x_0 \in \Omega$), then make several (finitely many) reflections from the graph of B_u, and its final velocity be not vertical, $v^+(t) \neq (0, 0, 1)$. At a point, say x_1, the x-projection $x(t)$ of the particle leaves Ω (this implies that $x_1 \in \partial\Omega$). Reparameterize the part of the trajectory between the point of the first impact and the point where the x-projection leaves Ω, the parameter being the path length s of the x-projection between x_0 and the current point.

Consider the z-coordinate $z(s)$ of the particle as a function of s, $0 \le s \le s_0$; here s_0 is the total length of the x-projection (which is a broken line) between x_0 and x_1; see Fig. 11. The breaks of the line correspond to the points of reflection of the particle.

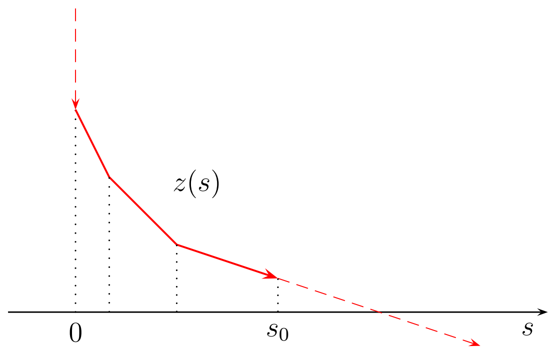

Figure 11: A particle trajectory parameterized by the path length. The breaks of the line correspond to reflections of the particle.

Note that $s_0 \ge \mathrm{dist}(x_0, \partial\Omega)$ and $0 \le z(s) \le M$ for all $0 \le s \le s_0$.

Let us show that the z-coordinate v_3 of the velocity v of the particle does not decrease at each impact. Indeed, for the velocities v and v^+ before and after the impact we have

$$v^+ = v - 2\langle v,\, n\rangle n,$$

and therefore, $v_3^+ = v_3 - 2\langle v,\, n\rangle n_3$, where n is the normal to $\partial(B_u)$ at the point of impact and $\langle v,\, n\rangle < 0$. There may be two cases:

(i) the particle reflects from the graph of u; then n is the upper normal to graph(u),

$$n = \frac{(-\nabla u(x),\ 1)}{\sqrt{1 + |\nabla u(x)|^2}};$$

(ii) the projection of the reflection point is a point of discontinuity of graph(u); that is, the particle is reflected from a "vertical wall". In this case the 3rd component of n is zero, $n = (n_1, n_2, 0)$.

In the case (i), $v_3^+ - v_3 > 0$; that is, the 3rd component of the velocity increases. In the case (i), $v_3^+ - v_3 = 0$; that is, the 3rd component of the velocity remains constant.

The following useful formula relates the derivative $z'(s)$ and the velocity v of the particle at the corresponding point:

$$\frac{(1,\ z'(s))}{\sqrt{1 + z'^2(s)}} = \left(\sqrt{v_1^2 + v_2^2},\ v_3\right).$$

It implies that the function $z'(s)$ is constant between impacts, and the increment of z' is nonnegative at each impact. Thus, the function $z(s)$ is convex, and therefore

$$z'(s_0) \geq \frac{z(s_0) - z(0)}{s_0} \geq -\frac{M}{d(x)}.$$

Recall the brief notation $d(x) = \text{dist}(x, \partial\Omega)$. Therefore the third component of the final velocity $v^+(x, u)$ of the particle satisfies

$$v_3^+(x, u) = \frac{z'(s_0)}{\sqrt{1 + z'^2(s_0)}} \geq -\frac{M/d(x)}{\sqrt{1 + M^2/d^2(x)}} = -\frac{M}{\sqrt{M^2 + d^2(x)}}. \qquad (23)$$

If, on the other hand, the final velocity is vertical then its 3rd component $v_3^+ = 1$ obviously satisfies (23).

Combining (4), (6), and (23), we obtain

$$F(u) \geq \phi(\Omega, M) \quad \text{for all} \quad u \in U_{\Omega,M}.$$

On the other hand, we have $S_{\Omega,M} \subset U_{\Omega,M}$, therefore by Theorem 1

$$\inf_{u \in U_{\Omega,M}} F(u) \leq \inf_{u \in S_{\Omega,M}} F(u) = \phi(\Omega, M).$$

Thereby the first relation in (9) is also proved.

3.2 Theorem 3

Since $D_{\Omega,M} \subset B_{\Omega,M}$, it suffices to prove the second equality in (10).

Let $\mathcal{N}_\varepsilon(A)$ denote the ε-neighborhood of the set A. Fix $\varepsilon > 0$ and denote

$$\tilde{\Omega} = \Omega \setminus \mathcal{N}_\varepsilon(\partial\Omega).$$

By Lemma 5 there exists a finite set of points $O_i \in \Omega \cap \mathcal{N}_\varepsilon(\partial\Omega)$, a domain $V \subset \tilde{\Omega}$ with $|V| < \varepsilon$, and a function $u \in S_{\tilde{\Omega},M/2}$ such that the particle corresponding to a point $x \in V$ is reflected vertically from graph(u), and the particle corresponding to a regular point of $\tilde{\Omega} \setminus V$ after the reflection passes through a point $(O_i, 0)$.

For each i find a ball $U_i \subset \mathbb{R}^2 \setminus \Omega$ such that the distance between O_i and each point of U_i is smaller than ε (see Fig. 12). In other words, $U_i \subset B_\varepsilon(O_i) \setminus \Omega$. Let K_i be the cone with vertex at $(O_i, M/2)$ and with the base

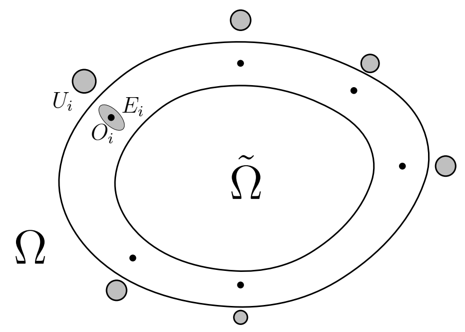

Figure 12: A part of the construction on the x-plane.

$U_i \times \{0\}$. That is, K_i is the union of rays with the endpoints at $(O_i, M/2)$ through all the points $(x, 0)$, $x \in U_i$. Let D_i be the intersection of K_i with $\partial\Omega \times [0,\ M/2]$.

Then for each i we choose $x_i \in U_i$ and denote by l_i the ray with the endpoint $(O_i, M/2)$ through $(x_i, 0)$. We select the points x_i in such a way that the resulting rays l_i do not mutually intersect.

Next for each i we choose a paraboloid of rotation P_i with focus at $(O_i, M/2)$, with the axis containing l_i, and such that $l_i \cap P_i = \emptyset$. Denote by $\hat{P}_i$ the convex hull of P_i; that is, $\hat{P}_i$ is the convex closed domain bounded by the paraboloid P_i. Our choice implies that the ray l_i is contained in the axis of P_i and in the domain $\hat{P}_i$. Denote $E_i = \hat{P}_i \cap (\mathbb{R}^2 \times \{M/2\})$. Thus, E_i is an ellipse (with its interior) containing O_i on the horizontal plane $z = M/2$ (see Fig. 12). We impose the additional conditions:

(i) the domains $\hat{P}_i$ are mutually disjoint;

(ii) $\hat{P}_i \cap (\partial\Omega \times \mathbb{R}) \subset D_i$;

(iii) $E_i \subset \Omega \setminus \overline{\tilde{\Omega}}$; that is, the ellipses E_i do not intersect $\partial\Omega$ and $\partial(\tilde{\Omega})$.

The conditions (i)–(iii) mean that the paraboloids P_i should be "sufficiently thin".

Take a body B_ε which is the union of four domains $B_i = B_i^\varepsilon,\ 1 \le i \le 4$,

$$B_\varepsilon = B_1 \cup B_2 \cup B_3 \cup B_4$$

(see Fig 13). Here B_1 is the subgraph of the function $M/2 + u$,

$$B_1 = \{(x, z) : x \in \overline{\tilde{\Omega}},\ 0 \le z \le M/2 + u(x)\},$$

and

$$B_2 = (\bar{\Omega} \times [0,\ M/2]) \setminus (\cup_i \hat{P}_i).$$

Further, we take two open domains $\Omega_0 \subset \Omega_1 \subset \Omega \setminus \tilde{\Omega}$ such that $\partial\Omega_1 \cap \partial\tilde{\Omega} = \emptyset$ and for each i, $\Omega_0 \cap E_i = \emptyset$ and $\overline{\cup_i E_i} \subset \Omega_1$. One can take, for instance, $\Omega_0 = \Omega \cap \mathcal{N}_{\varepsilon'}(\partial\Omega)$ and $\Omega_1 = \Omega \setminus \overline{\mathcal{N}_{\varepsilon'}(\tilde{\Omega})}$ with $\varepsilon' > 0$ sufficiently small.

We define

$$B_3 = \bar{\Omega}_0 \times [0,\ M],$$

$$B_4 = \bar{\Omega}_1 \times [M - \varepsilon',\ M],$$

where ε' is taken so small that the domain B_4 does not intersect $\mathrm{Conv}\big((\tilde{\Omega} \times [0,\ M]) \cup (\Omega \times [0,\ M/2])\big)$, and therefore is inaccessible for trajectories of the particles reflected from graph$(M/2 + u)$. The domain B_4 serves to "shield" the planar domains E_i (which are inlets of the "hollows" $\hat{P}_i \cap \{z \le M/2\}$) from incident particles with vertical direction. The domain B_3 serves to make the whole body B_ε connected.

We have $\bar{\Omega} \times \{0\} \subset B_\varepsilon \subset \bar{\Omega} \times [0,\ M]$. The surface $\partial B_\varepsilon \setminus (\partial\Omega \times \mathbb{R})$ is piecewise smooth by the definition of B_ε. As we will see below, the billiard scattering outside B_ε is regular. Therefore $B_\varepsilon \in B_{\Omega,M}$.

The domains $B_3 \cup B_4$ and B_1 are obviously connected. The section of B_2 by a horizontal plane $z = c,\ 0 \le c \le M/2$ is also connected and has nonempty intersection with B_1; therefore $B_1 \cup B_2$ is connected. Since the domains $B_3 \cup B_4$ and $B_1 \cup B_2$ have nonempty intersection, their union B_ε is connected.

If a flow particle incident on B_ε corresponds to a regular point of $\Omega \setminus \overline{\tilde{\Omega}}$, then either $x \in \Omega_1$, or $x \in (\Omega \setminus \overline{\tilde{\Omega}}) \setminus \Omega_1$. In both cases the particle is reflected vertically, and so, $v_3(x; B_\varepsilon) = 1$.

If a particle corresponds to a regular point of $\tilde{\Omega}$, then either $x \in V$, or x is a regular point of $\tilde{\Omega} \setminus V$. In the former case the particle is reflected vertically, and in the latter case the reflected particle passes through a point $(O_i, M/2)$ and then moves in $\hat{P}_i$. It may further happen that it makes one more reflection (which is necessarily from P_i) and then moves freely parallel to l_i (see Fig. 13).

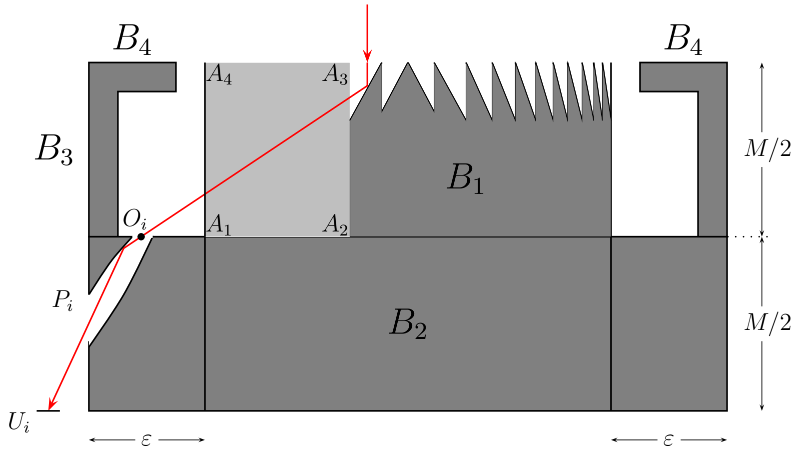

Figure 13: The cross section of the body B_ε by a vertical plane and the trajectory of a particle are shown. An unlikely case when all segments of the trajectory lie in one plane is depicted. The rectangle $A_1A_2A_3A_4$ shown lightgray projects on the valley V and therefore does not contain points of B_ε.

If the particle makes no reflections anymore, then it necessarily intersects $\partial\Omega \times [0,\, M/2]$ at a point of D_i, and then intersects the disc $U_i \times \{0\}$. In both cases the final motion is parallel to a line through $(O_i, M/2)$ and $(x, 0)$, $x \in U_i$. This implies that the 3rd component of the final velocity $v^+(x; B_\varepsilon)$ satisfies

$$v_3^+(x; B_\varepsilon) < -\frac{M/2}{\sqrt{M^2/4 + \varepsilon^2}}.$$

Thus, each particle makes no more than two reflections, and so, B_ε satisfies DIC. Since B_ε is connected, we conclude that $B_\varepsilon \in D_{\Omega,M}$.

The resistance of B_ε equals

$$R(B_\varepsilon) = \int_\Omega \frac{1 + v_3^+(x; B_\varepsilon)}{2}\, dx = \int_{(\Omega\setminus\tilde{\Omega})\cup V} (\cdots) + \int_{\tilde{\Omega}\setminus V} (\cdots)$$

$$< |\Omega \setminus \tilde{\Omega}| + |V| + |\tilde{\Omega} \setminus V| \cdot \frac{1}{2}\left(1 - \frac{M/2}{\sqrt{M^2/4 + \varepsilon^2}}\right).$$

Taking into account that $|\Omega \setminus \tilde{\Omega}| < \varepsilon|\partial\Omega|$, $|V| < \varepsilon$, and $|\tilde{\Omega} \setminus V| < |\Omega|$, we conclude that $R(B_\varepsilon) \to 0$ as $\varepsilon \to 0$. Theorem 3 is proved.

Acknowledgements

This work was supported by Portuguese funds through CIDMA – Center for Research and Development in Mathematics and Applications and FCT – Portuguese Foundation for Science and Technology, within the project PEst-OE/MAT/UI4106/2014, as well as by the FCT research project PTDC/MAT/113470/2009. The author is grateful to Evgeny Lakshtanov for the help in preparing Figure 2.

References

[1] A. S. Besicovitch. *The Kakeya problem.* Amer. Math. Monthly **70**, 697-706 (1963).

[2] F. Brock, V. Ferone and B. Kawohl. *A symmetry problem in the calculus of variations.* Calc. Var. **4**, 593-599 (1996).

[3] D. Bucur and G. Buttazzo. *Variational Methods in Shape Optimization Problems.* Birkhäuser (2005).

[4] G. Buttazzo. *A survey on the Newton problem of optimal profiles.* In "Variational Analysis and Aerospace Engineering", Springer Optim. Appl. **33**, Springer, New York, 33-48 (2009).

[5] G. Buttazzo, V. Ferone, and B. Kawohl. *Minimum problems over sets of concave functions and related questions.* Math. Nachr. **173**, 71-89 (1995).

[6] G. Buttazzo and B. Kawohl. *On Newton's problem of minimal resistance.* Math. Intell. **15**, 7-12 (1993).

[7] M. Comte, T. Lachand-Robert. *Newton's problem of the body of minimal resistance under a single-impact assumption.* Calc. Var. Partial Differ. Equ. **12**, 173-211 (2001).

[8] M. Comte, T. Lachand-Robert. *Existence of minimizers for Newton's problem of the body of minimal resistance under a single-impact assumption.* J. Anal. Math. **83**, 313-335 (2001).

[9] M. Comte and T. Lachand-Robert. *Functions and domains having minimal resistance under a single-impact assumption.* SIAM J. Math. Anal. **34**, 101-120 (2002).

[10] F. Cunningham. *The Kakeya problem for simply connected and for star-shaped sets.* Amer. Math. Monthly **78**, 114-129 (1971).

[11] C. Fefferman. *The multiplier problem for a ball.* Ann. Math. **94**, 330-336 (1971).

[12] B. Kawohl. *Some nonconvex shape optimization problems.* Optimal shape design (Tróia, 1998), 7-46, Lecture Notes in Math., **1740**, Springer, Berlin, 2000.

[13] T. Lachand-Robert and E. Oudet. *Minimizing within convex bodies using a convex hull method.* SIAM J. Optim. **16**, 368-379 (2006).

[14] T. Lachand-Robert, M. A. Peletier. *Newton's problem of the body of minimal resistance in the class of convex developable functions.* Math. Nachr. **226**, 153-176 (2001).

[15] I. Newton. *Philosophiae Naturalis Principia Mathematica.* 1687.

[16] A. Plakhov. *Newtons problem of minimal resistance for bodies containing a half-space.* J. Dynam. Control Syst. **10**, 247-251 (2004).

[17] A. Plakhov. *Precise solutions of the one-dimensional Monge-Kantorovich problem.* Sbornik: Math. **195**, 1291-1307 (2004).

[18] A. Plakhov. *Billiards and two-dimensional problems of optimal resistance.* Arch. Ration. Mech. Anal. **194**, 349-382 (2009).

[19] A Plakhov. *Comment on "Functions and domains having minimal resistance under a single-impact assumption" [SIAM J. Math. Anal. 34 (2002), pp. 101-120].* SIAM J. Math. Anal. **41**, 1721-1724 (2009).

[20] A Plakhov. *Optimal roughening of convex bodies.* Canad. J. Math. **64**, 1058-1074 (2012).

[21] A Plakhov. *Exterior Billiards: Systems with Impacts Outside Bounded Domains.* Springer, New York, 2012.

[22] A Plakhov. *The problem of minimal resistance for functions and domains.* SIAM J. Math. Anal. **46**, 27302742 (2014).

[23] G. Wachsmuth. *The numerical solution of Newtons problem of least resistance.* Preprint, Chemnitz University of Technology, 2013.