The minimum area of a simple polygon with given side lengths

K. Böröczky*, G. Kertész*, E. Makai, Jr.**

February 8, 2008

Abstract

Answering a question of H. Harborth, for any given $a_1, ..., a_n > 0$, satisfying $a_i < \sum_{j \neq i} a_j$ we determine the infimum of the areas of the simple n–gons in the Euclidean plane, having sides of length $a_1, ..., a_n$ (in some order). The infimum is attained (in limit) if the polygon degenerates into a certain kind of triangle, plus some parts of zero area. We show the same result for simple polygons on the sphere (of not too great length), and for simple polygons in the hyperbolic plane. Replacing simple n–gons by convex ones, we answer the analogous questions. The infimum is attained also here for degeneration into a certain kind of triangle.

Key words and phrases. Simple polygons, convex polygons, infimum of enclosed area, Euclidean plane, sphere, hyperbolic plane.

1991 Mathematics Subject Classification Primary:51M16; Secondary: 51M20, 51M25.

Research (partially) supported by Hungarian National Foundation for Scientific Research, Grant no. T002505, T016131 and T017314
** Research (partially) supported by Hungarian National Foundation for Scientific Research, Grant no. 41 and FKFP, Grant no. 0391/1997

1 Introduction

Let $a_1, ..., a_n > 0$, and let $a_i < \sum_{j \neq i} a_j, \ i = 1, ..., n$. It is well known that there is a unique simple n–gon in the Euclidean plane, with sides $a_1, ..., a_n$ (in this order) of maximal area, and this is the one inscribed into a circle (cf. e. g. [9], p. 44, taking into consideration [10], p. 57). Generalizations of this statement to arbitrary closed polygons, concerning the area of the convex hull, cf. in [13].

A special case of the dual question, i.e., of the infimum of the areas, has been posed by H. Harborth [6]: is it true, that for n odd, the area of a simple n–gon with unit sides is at least $\sqrt{3}/4$? As observed by [6], this can be attained in limit, cf. Fig. 1. Harborth, Kemnitz, Möller, Süssenbach [7] gave a positive answer to this question for $n = 5$. Our Theorem 1 will imply the positive answer to his question for each odd n.

Theorem 1 *For $n \geq 3$, the infimum of the areas of the simple polygons in the Euclidean plane, with side lengths $a_1, ..., a_n$ (in some order, depending on the polygon) — where $a_i \leq \sum_{j \neq i} a_j$, $i = 1, ..., n$ — equals min $\{A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)\}$, where $\{N_1, N_2, N_3\}$ is a partition of $\{1, ..., n\}$ with non-empty classes; $\varepsilon_i = \pm 1$; $\sum_{i \in N_j} \varepsilon_i a_i$, $j = 1, 2, 3$ are the side–lengths of a (possibly degenerate) triangle whose area is denoted by $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)$; and, if $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n) > 0$, then for $j = 1, 2, 3$ there is no partition $\{N'_j, N''_j\}$ of N_j such that both $\sum_{i \in N'_j} \varepsilon_i a_i$ and $\sum_{i \in N''_j} \varepsilon_i a_i$ are positive.*
If $n > 3$ then the infimum is never attained.

A simple polygon, having an area near to the infimum, is shown in Fig.2. The minimum appearing in Theorem 1 is well defined (it is defined on a non–void set), as it can be seen from the proof. Note that it is essential in Theorem 1 to assume that the polygon is simple (see 4.3.).

Apply Theorem 1 to the area A of a simple n–gon of unit sides where $n \geq 5$ is odd. Then a lower bound for A is provided by the infimum of the areas of triangles with integer edge lengths and odd perimeter, where for $j = 1, 2, 3$, N_j has no partition $\{N'_j, N''_j\}$ such that $\sum_{i \in N'_j} \varepsilon_i, \sum_{i \in N''_j} \varepsilon_i > 0$. The j-th side of such a triangle has length 1, and hence the area of this triangle is $\sqrt{3}/4$. In particular, $A > \sqrt{3}/4$.

The maximal area was attained for a convex polygon, so there was no separate question for convex polygons. However for the minimal area we have one, which is answered by our next theorem. Here we obtain a sharper result; namely, the order of the sides also can be fixed in advance.

Theorem 2 *For $n \geq 3$, the minimal area of a (not necessarily strictly) convex polygon in the Euclidean plane (degeneration to a doubly traversed segment admitted), with side lengths $a_1, ..., a_n$, in this order — where $a_i \leq \sum_{j \neq i} a_j$, $i = 1, ..., n$ —, equals* $\min\{A(N_1, N_2, N_3)\}$, *where $\{N_1, N_2, N_3\}$ is a partition of $\{1, ..., n\}$ with non-empty classes, of the form $\{\{n_1, ..., n_2 - 1\}, \{n_2, ..., n_3 - 1\}, \{n_3, ..., n_1 - 1\}\}$ (indices taken cyclically); $\sum_{i \in N_j} a_j$, $j = 1, 2, 3$ are the side lengths of a (possibly degenerate) triangle, whose area is denoted by $A(N_1, N_2, N_3)$. Moreover, the only convex polygons of the minimal area are those that degenerate to triangles (possibly degenerating to segments) of side-lengths $\sum_{i \in N_j} a_i$, $j = 1, 2, 3$, $\{N_1, N_2, N_3\}$ being a partition of $\{1, ..., n\}$ of the above form, with $A(N_1, N_2, N_3)$ minimal.*

The minimum appearing in Theorem 2, for $a_i \leq \sum_{j \neq i} a_j$, is well defined (it is defined on a non–void set). In fact, if it is well defined for some $n \geq 3$, then it is well defined for $n + 1$ as well. Namely let $a_1, ..., a_{n+1}$ be the cyclic order of sides of an $(n + 1)$–gon from Theorem 2. Then replace a_i and a_{i+1}, where $a_i + a_{i+1} = \min_j (a_j + a_{j+1})$ by $a_i + a_{i+1}$ (indices taken cyclically), thus obtaining the side lengths of a polygon with n sides, in this cyclic order. Then each side of this n–gon is at most the sum of its other sides. In fact, it only needs to be shown that the new side $a_i + a_{i+1}$ is not greater than the sum of the other sides, and this follows from $2(a_i + a_{i+1}) \leq \frac{1}{2} \sum_{j=1}^{n+1} (a_j + a_{j+1}) = \sum_{j=1}^{n+1} a_j$. Now a partition from Theorem 2 for this n–gon yields in a natural way a partition for Theorem 2 for our original $(n + 1)$–gon.

The same consideration shows that like in the Euclidean plane (see [11], p. 52-54), also in the hyperbolic plane, and on the unit sphere for $\sum a_i \leq 2\pi$, the inequalities $a_i \leq \sum_{j \neq i} a_j$ yield the existence of a (not necessarily strictly) convex polygon, possibly degenerating to a doubly traversed segment, with sides $a_1, ..., a_n$, in this order. If $a_i < \sum_{j \neq i} a_j$ then we even have convex polygons not degenerating to a segment; thus, simple polygons (see [11], p. 52-54 for the Euclidean case). Namely suppose we had above a polygon degenerating to a segment, with, say $a_1, ..., a_m$ covering it in one orientation, $a_{m+1}, ..., a_n$ in the other one, where necessarily $m, n - m \geq 2$. Then we can deform it to a non-degenerate convex quadrilateral with sides $a_1, a_2 + ... + a_m, a_{m+1} + ... + a_{n-1}, a_n$.

We mean convexity on the unit sphere in the strict sense (defined with the help of the unique connecting great circle arcs of length $< \pi$, that are required to exist for each pair of points). Then we have

Theorem 3 *The assertions of Theorems* 1 *and* 2 *hold for simple, or (not necessarily strictly) convex polygons on the unit sphere, of length* $\leq \pi$, *and in the hyperbolic plane (of arbitrary length), respectively.*

The paper is organized as follows. In Section 2 we deal with the Euclidean case, and prove Theorems 1 and 2. We consider the spherical and hyperbolic cases in Section 3, and we prove Theorem 3. Finally, we make some remarks and pose some related open problems in Section 4; among others we show a stability result related to Theorem 1 and the corresponding part of Theorem 3. Throughout the paper S^2 and H^2 denote the unit sphere and the hyperbolic plane, respectively.

2 The Euclidean case; proof of Theorems 1 and 2

We begin with the simple

Lemma 1 *Let a (possibly degenerate) triangle in the Euclidean plane have sides* a, b, x. *Then, for given* $a, b > 0$ *and for* $|a - b| \leq x \leq a + b$, *the area* A *of the triangle is a strictly concave function of* x.

Proof: Denoting by ξ the angle opposite to x, we have $A = \frac{1}{2}ab\sin\xi$ and it suffices to investigate $\sin\xi = \{1 - [(a^2 + b^2 - x^2)/(2ab)]^2\}^{1/2} := f(x)^{1/2}$. Here $f(x)^{1/2}$ is strictly concave if $f'(x)^2 - 2f(x)f''(x) > 0$, which is equivalent to the inequality

$$2(a^2 + b^2 - x^2)^2x^2 - [4a^2b^2 - (a^2 + b^2 - x^2)^2](a^2 + b^2 - 3x^2) > 0$$

for $|a - b| < x < a + b$. Let us suppose $a^2 + b^2 = 1$; then $C := 2ab \leq 1$. Set $1 - x^2 = u$, and hence we should verify $g(u) := 2u^2(1 - u) - (C^2 - u^2)(-2 + 3u) > 0$, where $|u| < C$. We have $g'(u) = 3(u^2 - C^2) < 0$ and $g(C) = 2C^2(1 - C) \geq 0$, therefore $g(u) > 0$ holds for $|u| < C$. Q.E.D.

Proof of Theorem 1:

1. First we show the weaker statement, where we ignore the requirement on the non–existence of the partitions $\{N_j', N_j''\}$ from the statement of the theorem. We apply induction on n. The case $n = 3$ being evident, from now on we suppose $n > 3$.

Let P be a simple polygon with sides $a_1, ..., a_n$ in this order. We have to show that the area $A(P)$ of P is at least the quantity given in the theorem, with the weakening from the last paragraph.

By Jordan's theorem on simple polygons (see [8]), a simple polygonal curve decomposes the plane to exactly two connected open regions, the interior and exterior regions. By [5], p. 157, Proposition 3 our simple polygon has a diagonal pq, entirely passing in the interior region. Let l denote the length of pq. The points p and q decompose the simple polygonal curve to two closed arcs $\widehat{pq}$, and adding the segment pq to any of these arcs we obtain a Jordan polygon again. These two Jordan polygons are denoted by P', P''. (For later reference we note that on S^2 we have that if the perimeter of P is at most π, then the perimeters of P' and P'' are strictly smaller than that of P.) We have $A(P) = A(P') + A(P'')$ as a consequence of Jordan's theorem, ([5], p. 157, Proposition 2).

Both P', P'' have at least 3, and at most $n - 1$ sides, so the induction hypothesis applies to them. Let P' or P'' contain the sides $a_1, ..., a_m$ or $a_{m+1}, ..., a_n$ of P, respectively. Then $A(P')$ is at least the area of a (possibly degenerate) triangle with sides $\varepsilon' l + \sum_{i \in M_1'} \varepsilon_i' a_i := \varepsilon' l + s_1'$, $\sum_{i \in M_2'} \varepsilon_i' a_i := s_2'$, $\sum_{i \in M_3'} \varepsilon_i' a_i := s_3'$, where $\{M_1', M_2', M_3'\}$ is a partition of $\{1, ..., m\}$ with $M_2', M_3' \neq \emptyset$ and $\varepsilon', \varepsilon_i' \in \{-1, 1\}$. We have the analogous statement for $A(P'')$, with $\{M_1'', M_2'', M_3''\}$ a partition of $\{m + 1, ..., n\}$ with $M_2'', M_3'' \neq \emptyset$ and $\varepsilon'', \varepsilon_i'' \in \{-1, 1\}$, and with s_1'', s_2'', s_3'' defined like above.

Let us denote the area of a (possibly degenerate) triangle with sides a, b, c by $A(a, b, c)$, where e.g. $c \geq |a - b| \geq 0$ holds by the triangle inequality. Then

$$A(P) = A(P') + A(P'') \geq A(\varepsilon' l + s_1', s_2', s_3') + A(\varepsilon'' l + s_1'', s_2'', s_3'') := F(l)\,.$$

Here $F(l)$ is a strictly concave function of l, because it is the sum of two strictly concave functions of l by Lemma 1. Its domain of definition is a (possibly degenerate) interval $[l_1, l_2]$, whose points l are characterized by the inequalities $|s_2' - s_3'| \leq \varepsilon' l + s_1' \leq s_2' + s_3'$, $|s_2'' - s_3''| \leq \varepsilon'' l + s_1'' \leq s_2'' + s_3''$. These triangle inequalities imply that all the six arguments $\varepsilon' l + s_1', ..., s_3''$ are non-negative. (For later reference we note that on S^2, if the perimeter of

P is at most π, then the perimeters of P', P'' are smaller than π, and thus $s_2' + s_3' < \pi,\ s_2'' + s_3'' < \pi$.)

The strict concavity yields that $F(l) \geq \min\{F(l_1), F(l_2)\}$, and for $l_1 < l < l_2$ even $F(l) > \min\{F(l_1), F(l_2)\}$. For $l \in \{l_1, l_2\}$ however we have $A(P'), A(P'') > 0$, and either $A(\varepsilon' l + s_1', s_2', s_3') = 0$ or $A(\varepsilon'' l + s_1'', s_2'', s_3'') = 0$. (For later reference we note that also in H^2, or on S^2, respectively, if equality $a = \pm b \pm c$ holds in the triangle inequality then the triangle has 0 area, where we assume $b + c < \pi$ on S^2.) Therefore in any case we have $A(P) > \min\{F(l_1), F(l_2)\}$.

Let us suppose that for $l_k \in \{l_1, l_2\}$ where $\min\{F(l_1), F(l_2)\}$ is attained, e.g. the triangle with sides $\varepsilon'' l_k + s_1'', s_2'', s_3''$ degenerates. Then we have $l_k = \varepsilon''(-s_1'' + \delta_2 s_2'' + \delta_3 s_3'')$, where $\delta_2, \delta_3 \in \{-1, 1\}$ with $\delta_2 + \delta_3 \neq -2$, and hence

$$\begin{aligned} A(P) &> \min\{F(l_1), F(l_2)\} = F(l_k) = A(\varepsilon' l_k + s_1', s_2', s_3') \\ &= A(-\varepsilon'\varepsilon'' s_1'' + \varepsilon'\varepsilon''\delta_2 s_2'' + \varepsilon'\varepsilon''\delta_3 s_3'' + s_1', s_2', s_3')\,. \end{aligned}$$

Here the last term is of the form $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)$, of the form required in the theorem, but with the weakening that we ignore the requirement about the non-existence of partitions $\{N_j', N_j''\}$.

2. In order to prove the theorem in its full strength, it is sufficient to show the following statement: if there is a partition $\{N_1, N_2, N_3\}$ of $\{1, ..., n\}$, and there exist signs $\varepsilon_1, ..., \varepsilon_n \in \{-1, 1\}$ fulfilling all requirements in the theorem, but the one about $\{N_j', N_j''\}$, then $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)$ cannot be minimal for all such partitions and signs satisfying the same requirements.

Let therefore $N_1, N_2, N_3, \varepsilon_1, ..., \varepsilon_n$ be as in the theorem, assume that $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n) > 0$ and there exists a partition $\{N_1', N_1''\}$ of N_1, with $s_1' := \sum\limits_{i \in N_1'} \varepsilon_i a_i,\ s_1'' := \sum\limits_{i \in N_1''} \varepsilon_i a_i > 0$ (thus $N_1', N_1'' \neq \emptyset$). Let us denote $s_j := \sum\limits_{i \in N_j} \varepsilon_i a_i,\ j = 2, 3$. Then $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n) = A(s_1' + s_1'', s_2, s_3) > 0$. Consider the triangle with sides $s_1' + s_1'', s_2, s_3$ as a simple quadrangle Q with sides s_1', s_1'', s_2, s_3, in this order, by inserting a vertex on the side $s_1' + s_1''$. The diagonal of Q, passing through the common endpoint of the sides s_1', s_1'' entirely passes in the interior of Q. Now we apply the results of part 1 of this proof, with the substitutions $P \to Q$, $n = 4$, $m = 2$, $a_1 \to s_3$, $a_2 = a_m \to s_1'$, $a_3 = a_{m+1} \to s_1''$, $a_4 = a_n \to s_2$, $\varepsilon' = \varepsilon_i' = \varepsilon'' = \varepsilon_i'' = 1$, $M_1' = M_1'' = \emptyset$, $M_2' = \{1\}$, $M_3' = \{2\}$, $M_2'' = \{3\}$, $M_3'' = \{4\}$, and thus $s_1' \to 0$, $s_1'' \to 0$, $s_2' \to s_3$, $s_3' \to s_1'$, $s_2'' \to s_1''$, $s_3'' \to s_2$. (For later reference we note that on S^2

the perimeter of Q is at most that of P, thus is at most π, if the same holds for P.) Then we gain, if e.g. the triangle with sides l_k, s_1'', s_2 degenerates, that $A(N_1, N_2, N_3; \varepsilon_1 ..., \varepsilon_n) = A(s_1' + s_1'', s_2, s_3) = A(Q) > A(\delta_2 s_1'' + \delta_3 s_2, s_3, s_1')$, with $\delta_2, \delta_3 \in \{-1, 1\}$. Here the last term is of the form required in the theorem, but without the requirement about $\{N_j', N_j''\}$. This shows our claim in part 2 of this proof, and this ends the proof of the theorem. Q.E.D.

Proof of Theorem 2: Let P be a polygon like in the theorem. If it degenerates to a segment or to a triangle, we are done. Let therefore P degenerate to a strictly convex k–gon Q, $4 \le k \le n$, with vertices $q_1, ..., q_k$, in this order. Denoting area by A, we have $A(P) = A(Q) = A(Q') + A(Q'')$, where $Q' = \text{conv}\{q_1, q_2, q_3, q_4\}$, $Q'' = \text{conv}\{q_4, q_5, ..., q_k, q_1\}$.

Let the diagonal q_1q_3 of the strictly convex quadrangle Q' have length l. Fixing the lengths of the sides of Q', we vary l, so that triangles $q_1q_2q_3$, $q_1q_4q_3$, if not degenerate, lie on different sides of q_1q_3. The obtained quadrangle $Q'(l)$ has an area $F(l) = A(\text{conv}\{q_1, q_2, q_3\}) + A(\text{conv}\{q_1, q_4, q_3\})$, which is by Lemma 1 a strictly concave function of l in the interval $[l_1, l_2] = [\max\{|q_1q_2 - q_3q_2|, |q_1q_4 - q_3q_4|\}, \min\{q_1q_2 + q_3q_2, q_1q_4 + q_3q_4\}]$. Here the original value l_0 of q_1q_3 lies in (l_1, l_2). Thus either $F(l)$ is strictly increasing in $[l_1, l_0]$ or strictly decreasing in $[l_0, l_2]$. Varying l, the angles of the triangles $q_1q_2q_3$, $q_1q_4q_3$ vary continuously, and in case one of these triangles degenerates, its angles tend to $\pi, 0, 0$ (π lying opposite to the longest side), or $0, \pi/2, \pi/2$ (0 lying opposite to the side q_1q_3), the latter case occurring exactly when the triangle is isosceles, with q_1q_3 as base, and $l \to 0$. (For later use we note that the same holds in H^2, and, provided that the perimeter of P, and thus of Q', is at most π, also on S^2. The very same remark applies to the strict concavity of $F(l)$.) For $k \ge 5$ let us place $Q'(l)$ to Q'', so that the vertices q_1 and q_4 of $Q'(l)$ lie in their original places and $Q'(l)$ lies on that side of q_1q_4 as Q'. Thus we obtain a closed k-gon $Q(l)$ (for $k = 4$ let $Q(l) = Q'(l)$), that is simple if $Q'(l)$ is convex quadrangle, possibly degenerating to a triangle, but not to a segment.

Now we vary l, beginning from $l = l_0$, either increasing or decreasing it, so that at increasing $|l - l_0|$ $F(l)$ strictly decreases. We stop at the first value l^*, where some angle of $Q(l)$ becomes π. (Below we will show that such l^* exists.) This angle must be at one of q_1, q_2, q_3, q_4. Note that for $k \ge 5$ and $Q'(l)$ not degenerating to a segment, the angle of $Q(l)$ at q_1, q_4 is greater than that of $Q'(l)$ at the same vertex.

If we have increased l, then, also considering the above remark on the limits of the angles (which we will also use below without further mentioning), for $l = l_2$ either $\angle q_1q_2q_3$, or $\angle q_1q_4q_3$ equals π (thus $q_1q_2q_3q_4$ degenerates to a triangle or segment), so the angle of $Q(l)$ at q_2 or q_4 is π, or at least π, respectively. Hence by continuity there exists a value $l^* \in (l_0, l_2]$ with the above asserted property.

If we have decreased l then for $l = l_1$ either $\angle q_1q_2q_3$ or $\angle q_1q_4q_3$ equals 0. For $l_1 = 0$, $Q'(l_1)$ degenerates to a convex deltoid degenerated to a segment, with angles π at q_1, q_3, thus again l^* exists. For $0 < l_1 = |q_1q_2 - q_3q_2| = |q_1q_4 - q_3q_4|$, $Q'(l_1)$ degenerates to a quadrangle degenerated to a segment, either with angles π at q_1 and q_3, or with angle 2π at q_1 or q_3. Hence in both cases l^* exists. For $0 < |q_1q_2 - q_3q_2| < |q_1q_4 - q_3q_4| = l_1$ or $0 < |q_1q_4 - q_3q_4| < |q_1q_2 - q_3q_2| = l_1$, $Q'(l)$ degenerates to a quadrangle degenerated to a triangle with one side (namely q_1q_3) elongated beyond one of its vertices. This quadrangle has an angle greater than π at the mentioned vertex. Thus again l^* exists.

If $k = 4$ then $Q(l^*)$ degenerated to a triangle or a segment, while for $k \geq 5$, $Q(l^*)$ degenerated to a simple polygon that is a strictly convex polygon with at most $k-1$, and at least 3 vertices (since $\operatorname{int} Q(l^*) \supset \operatorname{int} Q'' \neq \emptyset$). Moreover $A(Q) > A(Q(l^*))$. Repeating these considerations at most $n-3$ times, we obtain the statement of the theorem. Q.E.D.

3 The spherical and hyperbolic cases; proof of Theorem 3

We turn to the same questions on the unit sphere, and in the hyperbolic plane. (The question of maximal area cf. in [4], II. 2.1, II. 3.3, [15].)

For a simple polygon on S^2 there are two domains bounded by it, so we have two areas to be estimated. However, if the length of a closed polygonal line on S^2 is $< 2\pi$, then it lies in an open hemisphere ([4], II. 2.7). Therefore one of these areas is $< 2\pi$, the other one is $> 2\pi$. So for the lower estimate it suffices to investigate the domain lying in the above open hemisphere.

We begin with the analogue of Lemma 1 for the spherical and hyperbolic geometry. (For hyperbolic geometry cf. [1], [4], [12], [14]; in particular, the Caley-Klein model is exposed in [1], [12].)

Lemma 2 *Let a (possibly degenerate) triangle on the unit sphere or in the hyperbolic plane have sides a, b, x, where assume $a + b \leq \pi$ in the spherical case. Then, for $a, b > 0$ fixed and $|a - b| \leq x \leq a + b$, the area $\bar{A}$ of the* *triangle is a concave function of x. (For $x = a + b = \pi$ we define $\bar{A}$ by a limit procedure, i.e., we set $\bar{A} = \pi$.) In addition, the area is strictly concave in the hyperbolic case, and also in the sperical case if $a + b < \pi$.*

Proof:
1. First we consider the spherical case. Let the respective angles of the triangle be α, β, ξ. (For $x \to a + b = \pi$, and $a \le b$, say, fix the side b. Then the vertex with angle β moves on a small circle of spherical radius a, that is symmetrical with respect to the large circle containing the side b. Hence for $x \to a + b = \pi$, thus $\xi \to \pi$, we have $\bar{A} \to \pi$.) Then $\bar{A} = \alpha + \beta + \xi - \pi$, and $\cos x = \cos a \cos b + \sin a \sin b \cos \xi$, etc., hence for $|a - b| < x < a + b \le \pi$ we have

$$-\frac{d^2\bar{A}}{dx^2} = \frac{\cos\alpha}{\sin^3\alpha}\left(-\frac{\cos a}{\sin b}\cdot\frac{\cos x}{\sin^2 x} + \cot b\cdot\frac{1}{\sin^2 x}\right)^2 +$$
$$+\frac{1}{\sin\alpha}\left(\frac{\cos a}{\sin b}\cdot\frac{1+\cos^2 x}{\sin^3 x} - \cot b\cdot\frac{2\cos x}{\sin^3 x}\right) +$$
$$+\frac{\cos\beta}{\sin^3\beta}\left(-\frac{\cos b}{\sin a}\cdot\frac{\cos x}{\sin^2 x} + \cot a\cdot\frac{1}{\sin^2 x}\right)^2 +$$
$$+\frac{1}{\sin\beta}\left(\frac{\cos b}{\sin a}\cdot\frac{1+\cos^2 x}{\sin^3 x} - \cot a\cdot\frac{2\cos x}{\sin^3 x}\right) +$$
$$+\frac{\cos\xi}{\sin^3\xi}\cdot\frac{\sin^2 x}{\sin^2 a\sin^2 b} - \frac{1}{\sin\xi}\cdot\frac{\cos x}{\sin a\sin b}.$$

We prove that this expression is non–negative for $|a - b| < x < a + b \le \pi$, and is positive for $|a - b| < x < a + b < \pi$.

Let us multiply the first two terms of $-d^2\bar{A}/dx^2$ by $\sin\alpha/\sin a$, the second two terms by $\sin\beta/\sin b$, the last two terms by $\sin\xi/\sin x$ (these are equal and > 0). We can then eliminate from the obtained expression ξ, α, β, using

$$\frac{\cos\xi}{\sin^2\xi} = \frac{(\cos x - \cos a\cos b)\sin a\sin b}{\sin^2 a\sin^2 b - (\cos x - \cos a\cos b)^2},$$

and analogous equalities for α, β. Here all three such denominators equal $1 - \cos^2 a - \cos^2 b - \cos^2 x + 2\cos a\cos b\cos x > 0$. The expression obtained from $-d^2\bar{A}/dx^2$ by this multiplication and this elimination, $F(x)$, say, only contains trigonometric functions of a, b, x, and it suffices to show $F(x) \ge 0$, or $F(x) > 0$ in the respective domains, respectively.

Let $G(x) = F(x)(1 - \cos^2 a - \cos^2 b - \cos^2 x + 2\cos a \cos b \cos x) \sin a \sin b \sin^3 x$. It suffices to show $G(x) \geq 0$, or $G(x) > 0$, for $|a - b| < x < a + b \leq \pi$, or $|a - b| < x < a + b < \pi$, respectively. We will show these inequalities for the larger domains $0 \leq x \leq a + b$, where $0 \leq a, b$ and $a + b \leq \pi$, or $0 < x < a + b$, where $0 < a, b$ and $a + b < \pi$, respectively. We have

$$G(x) = (1 - X)\Big\{(A - BX)(B - AX)(A + B) + (X - AB)(1 - X^2)(1 + X)$$

$$+[(A + B)(1 - X) - X(1 + X)](1 - A^2 - B^2 - X^2 + 2ABX)\Big\}$$

$$:= (1 - X)H(x)\,,$$

where $A = \cos a$, $B = \cos b$, $X = \cos x$. Therefore it suffices to show $H(x) \geq 0$, or $H(x) > 0$ in the respective enlarged domains, respectively.

Let $A = E + F$ and $B = E - F$, and hence $E \leq 1$; and for $a, b \in [0, \pi]$, $a + b \leq \pi$ (or $a + b < \pi$) is equivalent to $E \geq 0$ (or $E > 0$, respectively). Then

$$\begin{aligned} H(x) &= E[2 - E - 2E^2 + (2 - E)X](1 - X)^2 + \\ &\quad +F^2[(1 - 6E) + (3 - 4E)X + (3 + 2E)X^2 + X^3]\,. \end{aligned}$$

Varying a, b in $[0, \pi]$, while fixing the value of E and X, $H(x)$ attains its minimum either for the minimum or for the maximum possible value of $|F|$.

If the coefficient of F^2 in $H(x)$ is non–negative, we decrease $|F|$ to 0; thus $H(x)$ does not increase. By what has been said above, by this decreasing $a + b \leq \pi$ (resp. $a + b < \pi$) will remain true, $0 \leq a, b, x$ (or $0 < a, b, x$) evidently remain true, and also $x \leq a + b$ (or $x < a + b$, respectively) remains true since for $0 \leq |F| \leq 1 - E$, $a + b = \arccos(E + F) + \arccos(E - F)$ decreases (not strictly) with increasing $|F|$. If the coefficient of F^2 in $H(x)$ is negative, we increase $|F|$, until either 1) $|F| = 1 - E$, i.e., $a = 0,\ b \geq 0$ or $b = 0,\ a \geq 0$, and in both cases $x \leq a + b$; or 2) $x = a + b$, and $|F| \leq 1 - E$, i.e., $a, b \geq 0$. Thus all inequalities $0 \leq a, b$, $a + b \leq \pi$, $0 \leq x \leq a + b$ remained true. Thus the proof of the inequality $H(x) \geq 0$ (or $H(x) > 0$, respectively) is reduced to the cases $a = b$, $b = 0$, say, and $x = a + b$.

First we settle the case $a = b$. Then $E = A = B \geq 0,\ F = 0$, further $0 \leq x \leq 2a \leq \pi$, thus $X \geq 2A^2 - 1$, and $H(x) = A[2 - A - 2A^2 + (2 - A)X](1 - X)^2$. Hence the second factor of the expression for $H(x)$ is $\geq 2A^2(1 - A) \geq 0$. Thus $H(x) \geq 0$, with equality if and only if $X = 1$, i.e., $x = 0$, or $E = A = 0$, i.e., $a = b = \pi/2$ (if $A = 1$ then $X = 1$, which case has been considered above).

Second we settle the case $b = 0$, i.e. $B = 1$. Then $H(x) = (-A + X)^3$, and we have $x \le a + b = a$, i.e. $X \ge A$. Hence $H(x) \ge 0$, with equality if and only if $x = a = a + b$.

Last we settle the case $x = a + b$. Then $1 - A^2 - B^2 - X^2 + 2ABX = \sin^2 a \sin^2 b - (\cos x - \cos a \cos b)^2 = 0$, $A - BX = \sin b \sin x$, $B - AX = \sin a \sin x$, $X - AB = -\sin a \sin b$. Thus

$$\begin{aligned} H(x) &= \sin a \sin b \sin^2 x \cdot (\cos a + \cos b - 1 - \cos x) \\ &= \sin a \sin b \sin^2(a+b) \cdot (\cos a + \cos b - \cos 0 - \cos(a+b)) \, . \end{aligned}$$

Here the last factor is non–negative, since if $0 \le t \le \pi/2$ and $0 \le s \le t$ then $\cos(t+s) + \cos(t-s)$ is a decreasing function of s (strictly decreasing for $t < \pi/2$). Hence $H(x) \ge 0$, with equality if and only if $a = 0$, or $b = 0$ or $a + b = \pi$.

Thus we have shown $H(x) \ge 0$, for $0 \le x \le a + b$, where $0 \le a, b$ and $a + b \le \pi$, thus also $-d^2\bar{A}/dx^2 \ge 0$ for $|a - b| < x < a + b$, where $0 < a, b$ and $a + b \le \pi$.

Now we show that actually $H(x) > 0$ for $0 < x < a+b$, where $0 < a, b$ and $a + b < \pi$. If the coefficient of F^2 in $H(x)$ is non–negative, then we have decreased $|F|$ to 0 (including the case that originally $F = 0$), thus not increasing $H(x)$, and this new value of $H(x)$ is positive, by $x > 0$ and $E > 0$. Therefore originally $H(x) > 0$ held. If the coefficient of F^2 in $H(x)$ is negative, then we have increased $|F|$ until we have equality in one of the inequalities $0 \le a$, $0 \le b$, $x \le a + b$ (and all these inequalities hold). Now originally none of these equalities can hold, thus $|F|$ has had to be strictly increased, thus $H(x)$ has been strictly decreased, and its new value is non–negative. Therefore originally $H(x) > 0$ held. All this shows that $-d^2\bar{A}/dx^2 > 0$ for $|a - b| < x < a + b$, where $a, b > 0$ and $a + b < \pi$.

2. The hyperbolic case is even simpler. With the same notation as above, we have $\bar{A} = \pi - \alpha - \beta - \xi$, $\cosh x = \cosh a \, \cosh b - \sinh a \, \sinh b \cos \xi$, etc., and for $0 < a, b$, $|a - b| < x < a + b$, $\sin \alpha / \sinh a = \sin \beta / \sinh b = \sin \xi / \sinh x > 0$. For $|a - b| < x < a + b$ one similarly calculates $-d^2\bar{A}/dx^2$, which is the same expression as $+d^2\bar{A}/dx^2$ in the spherical case, only here trigonometric functions of sides are replaced by hyperbolic ones, and the terms containing $1/\sin \alpha, 1/\sin \beta$ have positive signs in $-d^2\bar{A}/dx^2$. We are going to show that for $|a - b| < x < a + b$, where $0 < a, b$, we have $-d^2\bar{A}/dx^2 > 0$.

Analogously to the spherical case, multiplying $-d^2\bar{A}/dx^2$ by $\sin \alpha / \sinh a =$

$\sin\beta/\sinh b = \sin\xi/\sinh x$ and then eliminating α, β, ξ by using

$$\frac{\cos\xi}{\sin^2\xi} = \frac{(\cosh a\ \cosh b - \cosh x)\ \sinh a\ \sinh b}{\sinh^2 a\ \sinh^2 b - (\cosh x - \cosh a\ \cosh b)^2}$$

etc. (whose denominators equal $1-\cosh^2 a-\cosh^2 b-\cosh^2 x+2\cosh a\ \cosh b\ \cosh x > 0$) we obtain an expression $F(x)$. Now let $G(x) = F(x)(1-\cosh^2 a-\cosh^2 b - \cosh^2 x + 2\cosh a\ \cosh b\ \cosh x)\ \sinh a\ \sinh b\ \sinh^3 x$.

Again, these considerations reduce the problem to the inequality $G(x) \geq 0$ — where $G(x)$ is just the same polynomial of A, B, X as above, but now $A = \cosh a, B = \cosh b, X = \cosh x$ — for $0 \leq x \leq a + b$, where $0 \leq a, b$, or to the inequality $G(x) > 0$ for $0 < x < a + b$, where $0 < a, b$, respectively. (Actually the second inequality is sufficient for us, but we prove it through the first inequality.) We have $G(x) = (1 - X)H(x)$ and $X \geq 1$ (or $X > 1$), thus it suffices to show that $H(x) \leq 0$ (or $H(x) < 0$, respectively).

Defining E, F like above, we have $1 \leq E$, and similarly it suffices to investigate the cases $|F|$ is minimum or maximum possible, i.e., the cases $a = b$ (this obtained if the coefficient of F^2 in $H(x)$ is non–positive), $b = 0$ (say), $x = a + b$ (these obtained if the coefficient of F^2 in $H(x)$ is positive).

For $a = b$ we have $1 \leq X \leq 2A^2 - 1$ and $H(x) = A[2 - A - 2A^2 + (2 - A)X](1 - X)^2$. Hence the second factor of $H(x)$ is at most $2A^2(1 - A) \leq 0$ for $A \leq 2$, and at most $4 - 2A - 2A^2 < 0$ for $A \geq 2$. Thus $H(x) \leq 0$, with equality if and only if $x = 0$ (for $A = 1$ we have $X = 1$ that is included above).

For $b = 0$ we have $H(x) = (-A + X)^3$ and $X \leq A$, hence $H(x) \leq 0$, with equality if and only if $x = a = a + b$.

For $x = a+b$, similarly to the spherical case, we have $H(x) = \sinh a\ \sinh b\ \sinh^2(a + b)(\cosh a\ + \cosh b - \cosh 0 - \cosh(a + b)) \leq 0$, with equality if and only if $a = 0$ or $b = 0$.

Thus we have shown $H(x) \leq 0$ for $0 \leq x \leq a + b$, where $0 \leq a, b$, thus also $-d^2\bar{A}/dx^2 \geq 0$ for $|a - b| < x < a + b$, where $0 < a, b$.

The inequality $H(x) < 0$ for $0 < x < a + b$, where $0 < a, b$ follows from the above results like in the spherical case, hence actually $-d^2\bar{A}/dx^2 > 0$ for $|a - b| < x < a + b$, where $0 < a, b$. Q.E.D.

Proof of Theorem 3: The proofs of Theorems 1 and 2 carry over with some modifications. We have to use Lemma 2 in place of Lemma 1. We need yet the existence of a diagonal of our simple polygon, entirely passing in the interior region. (On S^2 we mean this so that our polygon lies in an open hemisphere, and we consider the region contained in that open hemisphere, that was sufficient to consider, cf. the remarks before Lemma 2.) This follows from the corresponding result on the Euclidean plane (cf. [5], p. 157), considering the central projection of the above open hemisphere to its tangent plane at its spherical centre, or using the Caley-Klein model of the hyperbolic plane, respectively. In both of these cases images of segments are segments (for S^2 the segment is supposed to lie on the open hemiphere), thus images of simple polygons are simple polygons, and the images of the interior regions are the interior regions of the images. On S^2 we have that the perimeter of our polygon is at most π. This ensures that at the applications of Lemma 2 the condition $a + b < \pi$ is satisfied. At the respective points we have referred to this fact, as well as to using some geometrical facts on S^2 and in H^2, in parentheses, in the proofs of Theorems 1 and 2. Q.E.D.

4 Concluding remarks

4.1. As can be seen from the proof of Theorem 1 (and the corresponding part of Theorem 3), we have proved a bit more:

Proposition 1 *Under the conditions of Theorem 1 (and the corresponding part of Theorem 3), if the order of the sides is fixed to be $a_1, ..., a_n$ (with $\sum a_i \leq \pi$ on S^2), then the area is at least the minimum given in Theorem 1 (and the corresponding part of Theorem 3), only restricting N_1, N_2, N_3 to consist of cyclically consecutive indices (like at Theorem 2); but then also N_j', N_j'' must consist of cyclically consecutive indices.*

Namely the same inductive step shows this too; and similarly for the statement about N_j', N_j''. Q.E.D.

However in this form the inequality is not sharp. For this we first show the following proposition, that is a stability variant of Proposition 1.

Proposition 2 *Let $n \geq 3$, $a_1, ..., a_n > 0$, $a_i < \sum\limits_{j \neq i} a_j$ (and $\sum a_i \leq \pi$ on S^2). Then the area of a simple n–gon P with sides $a_1, ..., a_n$, in this order, can approach the lower bound given in Proposition 1 arbitrarily close, only if it has a triangulation with the following properties. Each triangle in the triangulation, except one has a small area. The exceptional triangle has* *vertices at the end–points of the arcs I_j of* $\operatorname{bd} P$ *consisting of the sides a_i with $i \in N_j$, $j = 1, 2, 3$, with the distance of the end–points of I_j being near to some $\sum_{i \in N_j} \varepsilon_i a_i$. Here $\{N_1, N_2, N_3\}, \varepsilon_1, ..., \varepsilon_n$ are like in Proposition 1 and are such that they minimize $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)$ among all partitions and signs with the properties from Proposition 1.*

Proof: If $n = 3$ then the statement readily holds.

Let $n \geq 4$. We may assume that, at the inductive step at Theorem 1, for P' we chose M_1', M_2', M_3', ε', ε_i' so that $A(\varepsilon' l + s_1', s_2', s_3')$ was minimal among all such choices with the properties from Proposition 1, and similarly for P''. We recall the inductive step from Theorem 1:

$$A(P) = A(P') + A(P'') \geq A(\varepsilon' l + s_1', s_2', s_3') + A(\varepsilon'' l + s_1'', s_2'', s_3'') =$$

$$= F(l) \geq \min\{F(l_1), F(l_2)\} = F(l_k) = A(\varepsilon' l_k + s_1', s_2', s_3') =$$
$$= A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)\,.$$

Suppose that $A(P)$ approximates the lower bound given in Proposition 1 within η, where $\eta > 0$ is sufficiently small. Then $A(P) \leq A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n) + \eta$, since N_1, N_2, $N_3; \varepsilon_1, ..., \varepsilon_n$ is one of the systems for which the minimum could be taken in Proposition 1. (By part **2** of the proof of Theorem 1, the omission of the condition about N_j', N_j'' does not change the minimum of $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)$ in Proposition 1.) Further, $F(l) \leq F(l_k) + \eta$, thus, by strict concavity, l is near to l_k (for $F(l_1) = F(l_2)$ any of l_1, l_2 admitted as l_k), and therefore, e.g. $A(\varepsilon'' l + s_1'', s_2'', s_3'')$ is near to 0. Moreover we have $A(P') \leq A(\varepsilon' l + s_1', s_2', s_3') + \eta$ and $A(P'') \leq A(\varepsilon'' l + s_1'', s_2'', s_3'') + \eta$, hence $A(P'')$ is near to 0. Also we have (up to order of indices) $N_1 = M_1' \cup \{m+1, ..., n\}$, $N_2 = M_2'$, $N_3 = M_3'$.

By the assumed minimality property of $A(\varepsilon' l + s_1', s_2', s_3')$ and $A(\varepsilon'' l + s_1'', s_2'', s_3'')$ both P', P'' satisfy the hypothesis of the Proposition. By induction they satisfy the conclusion of the Proposition. (Note that there are only finitely many combinatorial possibilities for P', P'', so words like "near" can be used in a uniform sense for each of these possibilities.) The triangulations of P', P'' from the conclusion of the Proposition together give a triangulation of P, and each triangle in P'', and each but one triangle in P' has a small area. The exceptional triangle has vertices at the end–points of the arcs I_j' of $\operatorname{bd} P'$ (with analogous notation), hence at the end–points of the arcs I_j of $\operatorname{bd} P$. The distances of the end–points of I_j, $j = 2, 3$, are near to

$\sum_{i\in M'_j} \varepsilon'_i a_i = \sum_{i\in N_j} \varepsilon_i a_i$, while the distance of the end–points of I_1 is near to $\varepsilon' l + \sum_{i\in M'_1} \varepsilon'_i a_i$, thus is near to $\varepsilon' l_k + \sum_{i\in M'_1} \varepsilon'_i a_i = \varepsilon' l_k + s'_1 = \sum_{i\in N_1} \varepsilon_i a_i$. Finally $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_n)$ is minimal for all partitions and signs considered in the Proposition, since it is near to the area of the exceptional triangle, that is at most $A(P)$, and $A(P)$ approximates its lower bound in Proposition 1 within η. Q.E.D.

Proposition 3 *The inequality in Proposition 1, for fixed order of sides, is in general not sharp.*

Proof: We show our statement in R^2, but multiplying the sides of our example in R^2 by small positive numbers, we get examples in S^2, H^2, too.

Let us consider a simple 7-gon in R^2 whose consecutive sides $a_1, ..., a_7$ are, in this order, $1, x, 2x+1, x, y, 2y+1, y$, where x, y are integers, $1 << x << y$. Then $N_1 = \{1\}$, $N_2 = \{2,3,4\}$, $N_3 = \{5,6,7\}$, $\varepsilon_1 = \varepsilon_3 = \varepsilon_6 = 1$, $\varepsilon_2 = \varepsilon_4 = \varepsilon_5 = \varepsilon_7 = -1$ satisfy the properties from Proposition 1, and we have $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_7) = \sqrt{3}/4$.

We claim that any partition $\{N_1, N_2, N_3\}$ of $\{1, ..., 7\}$ with non–empty classes consisting of cyclically consecutive indices, and any signs $\varepsilon_1, ..., \varepsilon_7$, for which $\sum_{i\in N_j} \varepsilon_i a_i$ are the side lengths of a (possibly degenerate) triangle, satisfy $A(N_1, N_2, N_3; \varepsilon_1, ..., \varepsilon_7) \geq \text{const} > \sqrt{3}/4$, unless we have the above partition and above signs. Namely our triangle has integer sides, a, b, c say, and odd perimeter, hence e.g. $1 \leq c \leq a + b - 1$, $a + b + c \geq 3$. Now Heron's formula shows that the area is $\geq \sqrt{3}/4$, with equality if and only if $a = b = c = 1$, and otherwise the area is $\geq \sqrt{5}/4$. Suppose that $\max N_j \in \{3k-1, 3k, 3k+1\}$ $(k \in \{1,2\})$. In this case, if $N_j \not\supset \{3k-1, 3k, 3k+1\}$ or $\{3k-1, 3k, 3k+1\} \subset N_j$ but we do not have $\varepsilon_{3k-1} = -\varepsilon_{3k} = \varepsilon_{3k+1}$, then the triangle has a long side, and hence an area $\geq \sqrt{5}/4$. (Note that $N_j \cap \{3k-1, 3k, 3k+1\} = \{3k-1, 3k+1\}$ is impossible, since then $|N_j| = 6$, and then some $N_{j'}$ is empty.) If N_j strictly contains $\{3k-1, 3k, 3k+1\}$ and $k = 2$, then we can repeat the same reasoning for $N_j \setminus \{3k-1, 3k, 3k+1\}$, yielding that either the area is $\geq \sqrt{5}/4$, or N_j is the union of some classes $\{3k-1, 3k, 3k+1\}$ $(k \in \{1,2\})$ and $\{1\}$. Since each N_j is non-empty and $\sum_{i\in N_j} \varepsilon_i a_i \geq 0$, we get our claim.

Now apply Proposition 2. As the paragraph above shows, if the area approached the lower bound given in Proposition 1 arbitrarily close, then the exceptional triangle in Proposition 2 would have sides near to 1. This readily implies that the sides of lengths $2x+1, 2y+1$ would intersect, which is absurd. Q.E.D.

4.2. In order to prove Theorem 2 in R^2 only, a shorter proof can be given using Steiner's formula ([10], Theorem 14) for the area A of our convex quadrangle Q' from the proof of Theorem 2, rather than using our Lemma 1. Namely, the Steiner formula yields that $16A^2 = 4a^2b^2 + 4c^2d^2 - (a^2+b^2-c^2-d^2)^2 - 8abcd\cos(\alpha+\gamma)$, where a, b, c, d are the sides, $\alpha, \beta, \gamma, \delta$ are the angles of Q', in this order. We may suppose $\alpha + \gamma \geq \pi$; then like in Theorem 2 increase the length l of the diagonal separating the vertices with angles α, γ, thus increasing $\alpha + \gamma$ and decreasing A, till α or γ becomes π. This implies the existence of l^*, and the proof is finished like above.

4.3. If in Theorem 1, or in the corresponding statements of Theorem 3 we allow self–intersecting polygons, then for the definition of enclosed area, different from the signed area, we have e.g. the following possibilities. 1) The total area of all bounded connected components of the complement $R^2 \setminus P$ of the polygon P. 2) The sum of the integrals of the absolute values of the indices of P, taken for all bounded connected components of $R^2 \setminus P$. For both of these definitions the extension of Theorem 1, and of the corresponding statements of Theorem 3, are false. Like at Proposition 3, it suffices to give the common counterexample for R^2. The counterexample comes from perturbation of the oriented polygon $P = p_1p_2p_3p_4p_5$, where each edge has length 1, the points p_2 and p_5 coincide, and the segment $p_1p_2 = p_1p_5$ is orthogonal to the line p_3p_4 and lies outside of the regular triangle $p_2p_3p_4$. Now pull p_2 and p_5 slightly apart in a way that the edge lengths are kept and the edges p_2p_3 and p_4p_5 intersect, and the figure retains its original axis of symmetry. Then elementary calculations show that the area is decreased. Adding extra double unit edges results in a counterexample of unit sides for any odd $n \geq 7$. We note yet that if we do not want counterexamples of unit edges, then the above quadrangle $p_2p_3p_4p_5$ is already a counterexample. Adding still sides of length 0, or of length some small δ_0, we obtain counterexamples for any $n \geq 4$.

4.4. Presumably Theorem 1 and the corresponding parts of Theorem 3 hold, if we only assume that each point of R^2 (S^2, H^2), not on the oriented polygon, has an index 0 or 1 with respect to the oriented polygon (with area = the integral of the index). We ask if Theorems 1 and 2 hold for any Minkowski metric as well. It is still open whether the statement of Theorem

3 holds on S^2 if the length of the polygon lies in $(\pi, 2\pi]$.

4.5. We pose the similar question in R^d. If a polytope (finite union of non-degenerate d-simplices) has facet areas $A_1, ..., A_n$, then there is a (not necessarily strictly) convex polytope having the same facet areas and not smaller volume ([2], Theorem 2). Presumably for a convex polytope with fixed facet areas and (nearly) maximal volume this polytope must be in some sense near to a sphere (compare theorems in [4], II.4.3). Is the minimal volume of an arbitrary polytope with fixed facet areas $A_1, ..., A_n$ attained (in limit) if the polytope is a simplex plus some parts of zero volume, i.e., is the infimum of the volumes $\min\{V(N_1, ..., N_{d+1}; \varepsilon_1, ..., \varepsilon_n) | \{N_1, ..., N_{d+1}\}$ is a partition of $\{1, ..., n\}$ with non-empty classes; $\varepsilon_i = \pm 1$; $\sum_{i \in N_j} \varepsilon_i A_i$, $j = 1, ..., d+1$ are the facet areas of a (possibly degenerate) simplex$\}$ (where $V(N_1, ..., N_{d+1}; \varepsilon_1, ..., \varepsilon_n)$ is the infimum of the volumes of simplices with facet areas $\sum_{i \in N_j} \varepsilon_i A_i$, $j = 1, ..., d+1$)? One can pose the similar question for (not necessarily strictly) convex polytopes, but with $\varepsilon_i = 1$.

4.6. Last we sketch our original idea of proof of Theorem 1 and the corresponding part of Theorem 3; for simplicity of notation we restrict ourselves to R^2. It works on the level of heuristics, is considerably longer, but maybe the concepts in it have some interest. (Some of its details involving elementary plane topology have not been worked out, but most probably they could be.)

We used induction on n. Let $n \geq 4$. We wanted to move our n–gon P (i.e., deform it continuously with respect to a parameter varying in a non–degenerate closed interval of R) so that its side lengths, with their order, were fixed (say, this is $a_1, ..., a_n$), its area strictly decreased, and at the end of the motion (but not earlier) there occurred some double points of P (these cannot be inner points of some non-collinear sides both times). Moreover these double points should have been such that P touched itself "from the interior side", not "from the exterior side". Then the enclosed area is the sum of enclosed areas of some subcycles of P. E.g., we can take two subcycles P_1, P_2, one edge of P, a_1, say, being shared by P_1 and P_2, $a_2, ..., a_m$ belonging to P_1, $a_{m+1}, ..., a_n$ belonging to P_2. Then we have for the areas $A(P) = A(P_1) + A(P_2)$, and $A(P_i)$ can be estimated from below by induction, also using the length x of the edge of P_1 on a_1. Then the same concavity arguments work as in Section 2, giving the required lower estimate for $A(P)$.

For this inductive step we introduced the concept of a generalized simple oriented polygon P, defined by the properties that the winding number of x with respect to P is 0 or 1 for any $x \in R^2 \setminus P$, and for each oriented side $\overrightarrow{p_i p_{i+1}}$ of P, each point y in a neighbourhood of the open side $\overrightarrow{p_i p_{i+1}}$, lying in the exterior side of $\overrightarrow{p_i p_{i+1}}$ (with respect to the orientation) can be connected by an arc to infinity, avoiding P. We claimed Theorem 1 held for generalized simple (oriented) polygons P, with area = the integral of the index over R^2 (but with case of equality if and only if the set of points where the index is 1, either is an open triangle or is empty).

If the generalized simple polygon P was not actually simple, we had P_1, P_2 as above, and induction worked. If P was simple, we wanted to move P so that its area strictly decreased and at the end of the motion we obtained a non–simple generalized simple polygon of strictly smaller area, whose area already had been estimated from below. For this motion we used induction again. We considered $Q = \text{conv}P$; its vertices q_i (in this order) decompose P to arcs $\widehat{q_i q_{i+1}}$ (indices taken cyclically). For Q no triangle and no n–gon we moved the sides of P so that the arcs $\widehat{q_i q_{i+1}}$ were rigidly attached to the sides $q_i q_{i+1}$ of Q, while Q was moved according to the induction hypothesis, thus preserving the side lengths of P with their order, and strictly decreasing the area of Q, and hence that of P as well. We stopped at the first position when P ceased to be simple. For Q an n–gon we used arguments like in our Theorem 2 (except for $n = 4$ that had to be settled directly), while for Q a triangle we could enclose P in a simple (not strictly) concave quadrangle Q^* as well, with three vertices at the three vertices of Q, and fourth vertex some vertex of P, and use Q^* rather than Q. (The motion of Q^* is easily discussed.) All that remains to be shown is that the stopping position of P occurs at latest at the stopping position of Q (or Q^*, respectively), at the stopping position P becomes a generalized simple polygon (with orientation obtained from the positive orientation of its earlier positions, by passing to the limit), and that with P also P_1, P_2 from the beginning of this paragraph are generalized simple polygons (with the inherited orientations).

Acknowledgement: The authors express their gratitude to the Mathematical Institute of Oberwolfach for its hospitality during the Conference on discrete geometry in 1987, where this work had its origin. Further we express our sincere thanks to K. Böröczky, Jr., for his technical help in producing this paper.

References

[1] R. Baldus, F. Löbell: Nichteuklidische Geometrie, Sammlung Göschen 970, de Gruyter, Berlin, 1953

[2] I. Bárány, K. Böröczky, E. Makai, Jr., J. Pach: Maximal volume enclosed by plates and proof of the chessboard conjecture, Discr. Math., **60** (1986), 101-120

[3] L. Fejes Tóth: Lagerungen in der Ebene, auf der Kugel und im Raum, 2-te Auflage, Springer–Verlag, Berlin, 1972

[4] L. Fejes Tóth: Reguläre Figuren, Akad. Kiadó, Budapest, 1965

[5] B. Grünbaum: Polygons, The geometry of metric and linear spaces, East Lansing, 1974, L. M. Kelly ed., Lect. Notes in Math. 490, Springer–Verlag, Berlin, 1975

[6] H. Harborth: Oral communication, Oberwolfach conference on discrete geometry, 1987

[7] H. Harborth, A. Kemnitz, M. Möller, A. Süssenbach: Ganzzahlige planare Darstellungen der Platonischen Körper, Elem. Math. **42** (1987), 118-122

[8] W. Hurewicz, H. Wallman: Dimension theory, Princeton University Press, Princeton, N. J., 1941

[9] I. M. Jaglom, W. G. Boltjanski: Konvexe Figuren, Deutsch. Verl. Wiss., Berlin, 1956

[10] N. D. Kazarinoff: Geometric inequalities, Random House, New York–Toronto, 1961

[11] D. A. Kryžanovskiĭ: Isoperimetric figures. Maximum and minimum properties of geometrical figures, 3-rd ed., Gos. Izdat. Fiz.–Mat. Lit., Moscow, 1959 (in Russian)

[12] H. Liebmann: Nichteuklidische Geometrie, 2-te neubearb. Aufl., Göschen Verl., Berlin - Leipzig, 1912

[13] J. Pach: On an isoperimetric problem, Stud. Sci. Math. Hungar. **13** (1978), 43-45

[14] O. Perron: Nichteuklidische Elementargometrie der Ebene, Reihe Math. Leitfäden, Teubner, Stuttgart, 1962

[15] E. Schmidt: Die Brunn–Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie, I-II, Math. Nachr. **1** (1948), 81-157, **2** (1949), 171-244