EXHAUSTING AN AREA WITH DISCS

P. ERDÖS AND D. J. NEWMAN

ABSTRACT. The general question is, just how much of a given area may be covered by interior disjoint discs? We simplify matters by choosing the unit square as our given area, and then give estimates (upper and lower) on how much of it may be so covered by n discs.

It is true, although perhaps not trivial, that the unit square can have almost all its area exhausted by the removal of open disjoint discs in its interior. Our question is, how fast?

So define $R(n)$ to be the minimum of the areas one must leave behind when one removes n disjoint discs interior to S, the unit square. We have already remarked that $R(n) \to 0$ and our job is to estimate just how it does so. The straightforward estimates give upper and lower bounds which are enormously disparate ($e^{-c\sqrt{\log n}}$ and n^{-cn} respectively) and we try to repair this situation somewhat by showing that the truth is around n^{-B}. The precise result is

Theorem. $c_1 n^{-1/3} \geq R(n) \geq c_2 n^{-1}$, c_1 and c_2 positive constants.

Proof. We turn first to the lower bound and need the following elementary lemma.

Lemma 1. *Consider a disc cut by a chord of length c and suppose that the smaller region has height h and the larger one has height H. We then have*

(a) *The area of the smaller region is bounded by $(\pi/4)ch$.*

(b) *The area of the larger region is bounded by $(\pi/2)H^2$.*

To prove (a) consider the ellipse which has the chord as major axis and which is tangent to the minor arc at its midpoint. The ellipse and the circle meet at the endpoints of the chord and doubly at the tangency point and so they meet nowhere else (4 points are all they can share). At the endpoints the slope of the circle is smaller than ∞, the slope of the ellipse, and so the

entire circular segment must stay within the half ellipse. The area of the half ellipse is exactly $(\pi/4)ch$, however, and (a) is proved.

As for (b) simply construct the semicircle with center at midpoint of the chord and radius H. This is tangent to our original circle and hence contains the larger region. Its area is exactly $(\pi/2)H^2$, however, and (b) follows.

Now let E denote the strip, of height W, along the bottom of our square S, and consider any collection of disjoint discs in S which have their centers outside E. If D is any one of them then part (a) of the lemma insures that Area $(D \cap E) \leq (\pi/4)W \cdot$ (chord of D on the boundary of E). These chords are disjoint, however, and so the sum of these areas is bounded by $(\pi/4)W \cdot 1$.

Next look at a disc, D', in S with its center inside E. Here part (b) of our lemma applies and we conclude that Area$(D' \cap E) \leq (\pi/2)W^2$.

If we have n disjoint discs in S, then, the total amount of area they can cover in E is bounded by $(\pi/4)W + n \cdot (\pi/2)W^2$ so that there remains in E an uncovered area of $(1 - \pi/4)W - n(\pi/2)W^2$, choosing $W = (4 - \pi)/4\pi n$ shows that the uncovered area is at least $((4 - \pi)^2/32\pi) \cdot n^{-1}$ and our lower bound is established with $C_2 = (4 - \pi)^2/32\pi \approx .0074$.

As for our upper bound we remove n discs by the simple device of removing, at each turn, the largest disc which is contained in the residual set. (It is not quite trivial that this process even exhausts (almost all) the area of S but this is known and was perhaps first noticed by A. Beck.) We need the following

Lemma 2. *Suppose that a region has its boundary composed of 3 concave arcs (arcs whose every chord is disjoint from the region) and denote its area by a and perimeter by l. Suppose that a disc of radius r is inscribed therein (its interior lies in the region and its circumference meets all 3 arcs). Then $r \cdot l \geq 2a$.*

The simple proof is based on the integral formula $a = \frac{1}{2} \int_\gamma \vec{r} \cdot \vec{n} \, ds$ where γ is the boundary, $\vec{r}$ is the radius vector from the origin 0, which we take as the center of our disc, $\vec{n}$ is the unit outward normal and ds represents arc length.

If we focus on any one of our concave arcs A, and draw the two lines of support from 0 then the two points of contact define a subarc, A'. We observe first of all that for any point in $A - A'$, $\vec{r} \cdot \vec{n} \leq 0$. Secondly we note that, at any point of A', the line of support, L, separates the origin from A so that $\vec{r} \cdot \vec{n} =$ distance from 0 to $L \leq$ distance from 0 to $A = r$. In either case, then, $\vec{r} \cdot \vec{n} \leq r$ and so we have

$$a = \frac{1}{2} \int_\gamma \vec{r} \cdot \vec{n} \, ds \le \frac{r}{2} \int_\gamma ds = \frac{rl}{2}$$

as required.

This is clearly a best possible result as the example of the triangle shows. In our application, however, the three arcs will be mutually tangent circles and improvement is then possible. It can be shown e.g. that $r \cdot l \ge 2.8a$ and this would allow us to obtain $n^{-.4}$ rather than only $n^{-1/3}$. The calculations are very complicated however, and so we omit the details.

Now let us return to our strategy of removing the biggest disc possible. It is clear, then, that at the kth stage the residual set is composed of a finite number $(2k + 2)$ of disjoint "circular triangles" and that the next chosen disc will be one of their inscribed discs. Our lemma applies! Indeed if we call a_v and l_v the areas and perimeter of these circular triangles then we have $r_k l_v \ge 2a_v$ for all v (r_k being the radius of the next, or $(k + 1)$th, disc to be removed). Summing over v gives

(1) $r_k L_k \ge 2A_k$, where, of course, L_k and A_k are, respectively, the total perimeter and area of the residual set (the set remaining after the first k discs have been removed).

It is also clear, from the very definition of our strategy, that

(2). $r_{k-1} \ge r_k$,
(3) $L_{k-1} = L_k - 2\pi r_{k-1}$, and
(4) $A_{k-1} = A_k + \pi r_{k-1}^2$.

The trick is to notice that these entail the nonincrease of the quantity $\sqrt{r_k} L_k + 2A_k / \sqrt{r_k}$. We have, namely, by (3) and (4),

$$\left(\sqrt{r_{k-1}} L_{k-1} + 2A_{k-1}/\sqrt{r_{k-1}}\right) - \left(\sqrt{r_k} L_k + 2A_k/\sqrt{r_k}\right)$$

$$= \left(1/\sqrt{r_k} - 1/\sqrt{r_{k-1}}\right)\left(\sqrt{r_{k-1} r_k} L_k - 2A_k\right)$$

and by (2) the first factor is ≥ 0 while the second is $\ge r_k L_k - 2A_k$ which is, in turn, ≥ 0 by (1). The montonicity is thereby established and in particular we have

$$\sqrt{r_n} L_n + 2A_n/\sqrt{r_n} \le \sqrt{r_0} L_0 + 2A_0 / \sqrt{r_0} = 4\sqrt{2}.$$

Again, by (2) and (3), we obtain $L_n \ge n \cdot 2\pi r_n$ so that the above inequality becomes

$$r_n^{3/2} n\pi + r_n^{-1/2} A_n \le 2\sqrt{2}.$$

If, finally, we apply the arithmetic-geometric inequality to the left-hand side we find that it is

$$\ge 4[(r_n^{3/2}\pi n)(r_n^{-1/2}A_n/3)^3]^{1/4} = 4(3\pi n A_n^3)^{1/4}/3.$$

Combining gives $A_n \le 3/\sqrt[3]{4\pi n}$ and our result is established with $C_1 = 3/\sqrt[3]{4\pi} \approx 1.29$.

HUNGARIAN ACADEMY OF SCIENCES, BUDAPEST, HUNGARY

DEPARTMENT OF MATHEMATICS, YESHIVA UNIVERSITY, NEW YORK, NEW YORK 10033