ONE SQUARE AND AN ODD NUMBER OF TRIANGLES

OISÍN FLYNN-CONNOLLY

1. Introduction

Monsky's theorem states that a square cannot be dissected into an odd number of triangles. Despite an apparently simple statement, the proof quite complex, involving fairly high level mathematics. Two sections (parts 2 and 4) of this note thus provide a review of the standard results in valuation theory and combinatorics we will need. In part 3 we prove the existence of a valuation v such that $v(\frac{1}{2}) > 1$. In part 5 we use this valuation to construct a 3-colouring of the plane satisfying the axioms of Sperner's Lemma. Another property of this colouring, from which Monsky's theorem follows, is that any triangle with vertices having all different colours cannot have area $\frac{1}{n}$, for odd n. This note is an expanded rearrangement of the material in [1, Chapter 20] and Ronan O'Gorman's talk. The changes include more examples and the addition of a historical background.

2. Background to the problem

A well-researched class of questions in geometry and topology are the *dissection problems.* These ask if, given some geometric object (typically a polytope or ball), we can partition it into the finite pairwise disjoint union of smaller 'pieces.' Such a partitioning is known as a *dissection.* For example, given a some plane figure R, these pieces might be triangles of equal area. The resulting partition is referred to as an *equidissection.*

Such problems have been posed since antiquity. Euclid's celebrated proof of Pythagoras's Theorem rests on a triangle dissection. More recently, the Wallace-Bolyai-Gerwien theorem (1807) states that any two plane polygons of the same area may be decomposed into the same number of pairwise congruent triangles.

Another apparently simple question, and one that would have been easily understood by the ancient Greeks, is: for what $n \in \mathbb{N}$ can we dissect a square into n equally sized triangles?

So when Fred Richman of New Mexico State University was writing a Master's exam in 1965, he considered this problem a natural candidate for inclusion. It proved substantially less tractable than he originally expected, and he was surprised to be unable to find a proof or even a reference for it. Eventually he wrote to the *American Mathematical Monthly* to pose it as a challenge to other researchers.

Five years passed. It wasn't until 1970, Paul Monsky - building upon the work of John Thomas (who proved it when all triangles have rational coordinates) - was finally able to publish a full solution. His proof, which will be the object to this report, was seemingly shooting cannons at sparrows - utilizing both combinatorial

---

Date: October 30, 2018.

topology and valuation theory. Perhaps the biggest surprise of all is that, nearly 50 years later, this remains the only known proof.

3. The theory of valuations

Definition 3.1. Given some field $\mathbb{K}$, *a valuation on* $\mathbb{K}$ *to* $\mathbb{R}_{\geq 0}$ is a map

$$v : \mathbb{K} \to \mathbb{R}_{\geq 0}$$

satisfying the following:

(1) $v(x) = 0$ if and only $x = 0$.
(2) $v(xy) = v(x)v(y)$ for all $x, y \in \mathbb{K}$.
(3) $v(x + y) \leq v(x) + v(y)$ for all $x, y \in \mathbb{K}$.

Example 3.2. The absolute value function $\| \cdot \|$ on $\mathbb{R}$. Indeed, generalizing $\| \cdot \|$ was the original motivation for valuations.

Example 3.3. Every field has the *trivial valuation*. This is defined as

- $v(x) = 1$ if $x \in \mathbb{K} \backslash \{0\}$.
- $v(0) = 0$.

Example 3.4. If v is a valuation on some field, then v^t is also a valuation for all $t \in \mathbb{R}^+$.

Definition 3.5. A valuation on $\mathbb{K}$ is called *non-Archimedean* if $v(x+y) \leq \max(v(x), v(y))$ for all $x, y \in \mathbb{K}$.

Remark 3.6. $v(x)$ and $v(y)$ are non-negative by condition (1) in the definition of valuations. Therefore $v(x + y) \leq \max(v(x), v(y)) \leq v(x) + v(y)$. We conclude that the non-Archimedean condition is nothing but a stronger version of condition (3).

Example 3.7. Let p be a prime number. Then given nonzero $r \in \mathbb{Q}$, r is of the form $p^k \cdot \frac{a}{b}$ where $\gcd(ab, p) = 1$. We define the *p-adic valuation on* $\mathbb{Q}$ as $\|r\|_p = p^{-k}$ and $\|0\|_p = 0$.

Lemma 3.8. *$\| \cdot \|_p$ is a non-Archimedean valuation on $\mathbb{Q}$.*

Proof. It is obvious that $\|\cdot\|_p$ satisfies the first two defining conditions of a valuation. Now let $r = p^k \frac{a}{b}$ and $s = p^l \frac{c}{d}$ where we assume without loss of generality that $k \geq l$. In other words, $\|r\|_p \leq \|s\|_p$. Now

$$\|r + s\|_p = \|p^k \frac{a}{b} + p^l \frac{c}{d}\|_p \leq p^{-l} \|\frac{p^{k-l}ad + bc}{bd}\|_p \leq p^{-l} \leq \max(\|r\|_p, \|s\|_p).$$

So $\| \cdot \|_p$ satisfies the non-Archimedean condition. This is stronger than (3) so we can conclude $\| \cdot \|_p$ is a non-Archimedean valuation on $\mathbb{Q}$. □

One of the reasons p-adic valuations are of interest is that all non-trivial valuations on $\mathbb{Q}$ are either p-adic or nearly so. The following theorem of Ostrowski, which we shall not prove, makes this concrete.

Theorem 3.9. *Every nontrivial valuation on $\mathbb{Q}$ is either $\| \cdot \|_p$ for a unique prime p, the standard absolute value on $\mathbb{Q}$ or $\| \cdot \|_p^s$ for a positive real number s and a unique prime p.*

The notion of a valuation as given in Definition 3.1 is insufficiently general for our aims. To remedy this we will first need to introduce the idea of an ordered abelian group.

Definition 3.10. *An ordered abelian group (OAG)* is an abelian group $(G, \cdot)$ equipped with a total order $<$ such that

$$a < b \implies a \cdot c < b \cdot c \text{ for all } c \in G.$$

We usually consider an ordered abelian group together with $0 \notin G$ such that

- $0 \cdot a = 0$ for all $a \in \{0\} \cup G$.
- $0 < a$ for all $a \in G$.

Example 3.11. $\mathbb{R}^+$ under multiplication with the usual ordering $<$ is an OAG.

As we are taking the operation on ordered abelian groups to be multiplication, we will refer to the identity element as 1.

Remark 3.12. Suppose a group G with total ordering $<'$ has torsion i.e. an element $g \neq 1$ such that $g^n = 1$. As $\neq$ is a total order we have either $g \neq 1$ or $1 \neq g$. If $g \neq 1$ we obtain $g <' g^2$. and, continuing, $1 <' g <' g^2 <' \cdots <' g^n = 1$. This gives $1 \neq 1$, a contradiction to the non-reflexivity of $\neq$. An symmetric argument gives a contradiction when $1 \neq g$. We conclude that any OAG must be torsion-free.

Conversely, it can be shown any torsion-free abelian group G admits a total order $<$ such that $(G, \cdot, <)$ is an OAG.

Definition 3.13. Given some field $\mathbb{K}$ and an OAG G, *a valuation on* $\mathbb{K}$ *with values in* G is a map

$$v : \mathbb{K} \to \{0\} \cup G$$

satisfying the following:

(1) $v(x) = 0$ if and only $x = 0$.
(2) $v(xy) = v(x)v(y)$ for all $x, y \in \mathbb{K}$.
(3) $v(x + y) \leq v(x) + v(y)$ for all $x, y \in \mathbb{K}$.

v is called non-Archimedean if $v(x + y) \leq \max(v(x), v(y))$ for all $x, y \in \mathbb{K}$.

We will now prove a lemma establishing the basic properties of valuations.

Lemma 3.14. *Let v be a valuation from $\mathbb{K}$ with values in G. Then we have*

(1) $v(1) = 1$.
(2) $v(-x) = v(x)$ for all $x \in \mathbb{K}$.
(3) $v(x^{-1}) = v(x)^{-1}$ for all $x \in \mathbb{K}$.
(4) If $v(x) < 1$ and $x \neq 0$ then $v(x^{-1}) > 1$
(5) If v is non-Archimedean and we have $x, y \in \mathbb{K}$ such that $v(x) > v(y)$ then $v(x + y) = v(x)$.
(6) If v is non-Archimedean and we have $x, y_1, y_2 \ldots, y_n \in \mathbb{K}$ such that $v(x) > v(y_1), v(y_2), \ldots, v(y_n)$ then $v(x + y_1 + y_2 + \cdots + y_n) = v(x)$.

Proof.

(1) $v(1) = v(1 \cdot 1) = v(1) \cdot v(1)$. Therefore we have $v(1) = 1$.
(2) $v(-1)^2 = v(-1 \cdot -1) = v(1) = 1$. G is torsion-free and in particular has no elements of order 2. Thus $v(-1) = 1$. Then $v(-x) = v(-1) \cdot v(x) = v(x)$.
(3) $1 = v(x \cdot x^{-1}) = v(x) \cdot v(x^{-1}) \implies v(x^{-1}) = v(x)^{-1}$.
(4) By (2), $v(x^{-1}) = v(x)^{-1}$. Now $v(x) > 1 \implies 1 > v(x)^{-1}$ as required.

(5) $v(x) = v(x+y+(-y)) \leq \max(v(x+y), v(-y)) \leq \max(v(x+y), v(y)) \leq \max(\max(v(x), v(y)), v(y)) = v(x)$. The two ends of this chain of inequalities are equal, and thus all these inequalities become equalities. In particular $v(x) = \max(v(x+y), v(y))$ and $v(x) \neq v(y)$ by assumption, so $v(x+y) = v(x)$.

(6) We use induction on n. When $n = 1$ we have $v(x) = v(x + y_1)$ by (5). Assume the inductive hypothesis for $n = k$. Then $v(x+y_1+y_2+\cdots+y_k) = v(x) > v(y_{k+1})$. Thus, by (5) we have $v(x+y_1+y_2+\cdots+y_{k+1}) = v((x+y_1+y_2+\cdots+y_k)+y_{k+1}) = v(x+y_1+y_2+\cdots+y_k) = v(x)$ as required. □

The following definition and lemma are, up to the direction of inequalities, standard in commutative algebra.

Definition 3.15. Let v be a valuation on $\mathbb{K}$ with values in some OAG. Then

$$R = \{x \in \mathbb{K} : v(x) \leq 1\}$$

is a subring of $\mathbb{K}$, called the *valuation ring* corresponding to v.

Lemma 3.16. *Given a field $\mathbb{K}$ and a proper subring R, let $R^{-1} := \{x \in \mathbb{K} : x^{-1} \in R, x \neq 0\}$. Then there exists a valuation v from $\mathbb{K}$ into some OAG such that R is the valuation ring corresponding to v if and only if $\mathbb{K} = R \cup R^{-1}$.*

Proof. Suppose we have our valuation v and R is the valuation ring corresponding to v. Suppose $v(x) \notin R. \implies v(x) > 1$. Then by Lemma 3.14, $v(x^{-1}) < 1 \implies x^{-1} \in R$. So $R \cup R^{-1} = \mathbb{K}$.

Now suppose $R \cup R^{-1} = \mathbb{K}$. Let $U := \{x \in R : x \text{ invertible in } R\}$. Also let

$$G := \mathbb{K}^{\times}/U.$$

We turn G into an ordered abelian group by imposing the following order relation on it.

$$xU < yU \Leftrightarrow xy^{-1} \in R \backslash U.$$

It is straightforward to verify that this satisfies the order axioms. Define:

$$v(x) := xU.$$

It is easy to see v is a valuation and that R is the valuation ring with respect to it. □

And thus concludes our brief introduction to valuations. In the next subsection, we will draw upon much of this theory to prove a highly non-trivial result on the existence of a valuation on $\mathbb{R}$ with certain prescribed properties.

4. A valuation with $v(\frac{1}{2}) > 1$

Theorem 4.1. *There exists a non-Archimedean valuation v from $\mathbb{R}$ to some ordered abelian group satisfying $v(\frac{1}{2}) > 1$.*

We will give two proofs. The first of these is at least partially constructive and will hinge on a powerful result from algebra. The second is that given explicitly in *Proofs from THE BOOK.*

Proof 1. Consider the 2-adic valuation on $\mathbb{Q}$. Now $\|\frac{1}{2}\|_2 = 2^1 = 2 > 1$.

Chevalley's theorem from algebra (which's proof we shall omit) states:

Theorem. *Any non-Archimedean valuation v of a field F into an algebraically ordered ring Φ with divisible multiplicative group can be extended to a valuation of every field E containing F.*

We have previously shown that the 2-adic valuation is non-Archimedean. Taking F to be $\mathbb{R}$ and E to be $\mathbb{Q}$, we see that the 2-adic valuation can be extended from $\mathbb{Q}$ to $\mathbb{R}$ and this extension has the requisite properties. □

Remark. Note that we have actually proven something much stronger than Theorem 4.1. We have shown the existence of a valuation from from $\mathbb{R}$ to $\mathbb{R}_{\geq 0}$, that is, a valuation in the classical sense.

Proof 2. Consider the set M of subrings $R \subseteq \mathbb{R}$, $\frac{1}{2} \notin R$ ordered by inclusion. Every ascending chain of subrings

$$B_1 \subset B_2 \subset \cdots \subset B_i \subset \cdots$$

is bounded above by $\cup_{i \in \mathbb{N}} B_i$. M thus satisfies the conditions of Zorn's Lemma and has a maximal element. Therefore there exists an inclusion-maximal subring $R \subseteq \mathbb{R}$ satisfying $\frac{1}{2} \notin R$.

Next we prove the following lemma:

Lemma 4.2. *Any inclusion-maximal subring $R \subseteq \mathbb{R}$ satisfying $\frac{1}{2} \notin R$, is a valuation ring.*

Proof. Assume R is not a valuation ring. Then, by Lemma 3.16, there exists an element $a \in \mathbb{R}$ which is not in R or R^{-1}. Now consider $R[a]$ and $R[a^{-1}]$. Both of these contain R, which was the maximal subring not containing $\frac{1}{2}$. Thus both contain $\frac{1}{2}$. Therefore $2R[a] = R[a]$ and $2R[a^{-1}] = R[a^{-1}]$. In particular, $2R[a]$ and $2R[a^{-1}]$ both contain 1. So we can write:

$$1 = 2u_0 + 2u_1 a + 2u_2 a^2 + \cdots + 2u_m a^m \text{ with } u_i \in R \tag{1}$$

$$1 = 2v_0 + 2v_1 a^{-1} + 2u_2 a^{-2} + \cdots + 2v_n a^{-n} \text{ with } v_i \in R \tag{2}$$

where we have chosen the representations such that n and m are minimal. We assume without loss of generality that $m \geq n$.

We multiply equation (2) by a^n and subtract $2v_0 a^n$ from both sides to obtain

$$(1 - 2v_0)a^n = 2v_1 a^{n-1} + 2u_2 a^{n-2} + \cdots + 2v_n \tag{3}$$

Now multiply (1) by $1 - 2v_0$ and add $2v_0$ giving

$$1 = 2(u_0(1 - 2v_0) + v_0) + 2u_1(1 - 2v_0)a + \cdots + 2u_m(1 - 2v_0)a^m. \tag{4}$$

We can substitute here for the term $(1 - 2v_0)a^m$, by taking equation (3) and multiplying it by a^{m-n}. We obtain an expression for 1 where a is of degree strictly less than m. This contradicts our earlier assumption – thus proving the lemma. □

By the lemma, the maximal subring existing by virtue of Zorn's Lemma is a valuation ring. Therefore there exists a valuation ring R such that $\frac{1}{2} \notin R$. We consider the valuation v corresponding to R. Note that $\frac{1}{2} \notin R \implies v(\frac{1}{2}) > 1$. So we can simply take v to be the valuation in the statement of the theorem. □

The following property of v will prove useful later.

Lemma 4.3. *For v in the statement of Theorem 4.1 and n an odd integer, we have $v(\pm\frac{1}{n}) = 1$.*

Proof. $v(\frac{1}{2}) > 1 \implies v(2) < 1$ Then $v(2+2) \le max v(2), v(2) = v(2) < 1$ and, by induction, $v(2k) < 1$ for all positive integers k. $v(2k+1) = v(1) = 1$ by lemma 3.14 part (5). Thus $v(\frac{1}{2k+1}) = v(-\frac{1}{2k+1}) = 1$. as required. □

5. Sperner's Lemma

Given the statement of the problem, if Monsky's proof drew only from the world of valuations it would be strange enough, but we are about to take yet another esoteric detour.

Brouwer's fixed-point theorem is a well known result in topology stating that for any continuous function f mapping a compact convex set – usually taken to be the n-ball – to itself there is a point x_0 such that $f(x_0) = x_0$. Less well-known perhaps is the theorem's combinatorial analog, to which it is equivalent – Sperner's Lemma. In full generality this states:

Theorem 5.1. *Let $\mathcal{A}$ be to a n-dimensional simplex $A_1A_2 \ldots A_{n+1}$. We consider a triangulation T which is a disjoint division of $\mathcal{A}$ into smaller n-dimensional simplices. Denote the colouring function as $f : S \to \{1, 2, 3, \ldots, n, n+1\}$, where S is again the set of vertices of T. The rules of colouring are:*

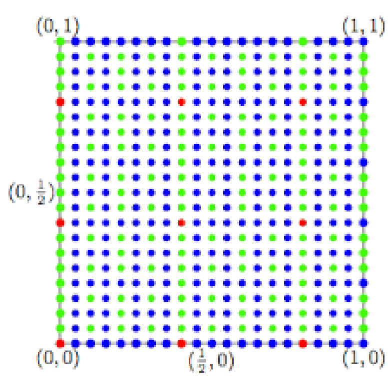
- *The vertices of the large simplex are coloured with different colours.*
- *Vertices of T located on any k-dimensional subface of the large simplex $A_{i_1}A_{i_2} \ldots A_{i_{k+1}}$ are colored only with the colours $i_1, i_2, \ldots, i_{k+1}$*

Then there exists an odd number of simplices from T, whose vertices are coloured with all $n+1$ colours. In particular, there must be at least one.

We will be interested the following very specific version of Sperner's Lemma specialized to two dimensions.

Theorem 5.2. *Colour $\mathbb{R}^2$ with three colours, with a colouring such that every line of the plane receives at most two colours. We require also that $(0,0), (0,1)$ and $(1,0)$ are different colours. We further require $(0,0)$ is not the same colour as $(1,1)$. Every dissection of the unit square $S = [0,1]^2$ into finitely many triangles contains an odd number of triangles with vertices all different colours (BGR triangles), and thus at least one.*

The figure below shows the colour for each point in the unit square whose coordinates are fractions of the form $\frac{k}{20}$, for one such colouring of the plane. The particular colouring chosen will be important in the next section.



We shall give the traditional proof of this result, attributed to a certain Emanuel Sperner.

Proof. We will take the three colours to be red, blue and green. Given a dissection of $S = [0,1]^2$, consider the segments between each vertex. We will call a segment a *red-blue segment* if one endpoint is red and the other is blue. We can assume without loss of generality that $(0,0)$ is red, $(1,0)$ is blue and thus $(0,1)$ is green.

First, observe that bottom edge of the square must contain an odd number of red-blue segments. This follows from the fact $(0,0)$ is red and $(1,0)$ is blue, by assumption the edge has only red and blue vertices, and therefore the walk from one end to another must have an odd number of changes between red and blue. Also note that the other edges of the square contain no red-blue segments. Therefore the edge of the square contains an odd number of red-blue segments.

Secondly, observe that for any triangle $\triangle$ in the dissection with at most two colours on its vertices, an even number of its edges, either two or zero, are red-blue segments. Every BGR triangle however, has one red-blue segment on its boundary – an odd number.

Lastly, we count the number of red-blue edges summed across every triangle in the dissection. This can be accomplished in two ways. First, counting each triangle is equivalent to counting every segment in the interior of the square twice and those on the boundary once. Since we know the bottom edge of the square contains an odd number of red-blue segments our total is odd. Then we can count the individual triangles directly. By our second observation, each BGR triangle contributes an odd number and every other triangle gives an even number. Thus we have an odd number of BGR triangles and, in particular, at least one. $\square$

Remark. The first observation – that bottom edge of the square contains an odd number of red-blue segments – is actually just Sperner's Lemma in one direction. This hints that one can adapt this proof to the general case by induction on the dimension .

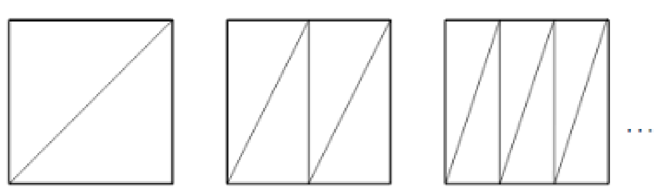
We are now finally ready to meet our main result.

6. Monsky's Theorem

Theorem 6.1. *It is possible to dissect a square into an n triangles of equal area if and only if n is even.*

The proof we will give is slightly simplified version of Monsky's original colouring argument. The simplification is due to Hendrik Lenstra.

Proof. One direction of the proof is almost trivial. If n is even we can divide both the top and bottom edge of the square into $\frac{n}{2}$ equal intervals and connect the endpoints in a zigzag fashion as illustrated below.



If n is odd, things become more involved. By translation and homothety we can assume the square in question is $[0,1]^2$ By Theorem 4.1, there exists a valuation v from $\mathbb{R}$ to some ordered abelian group satisfying $v(\frac{1}{2}) > 1$. Use v to construct a colouring $f : \mathbb{R}^2 \to \{\text{blue}, \text{green}, \text{red}\}$ as follows. Given $(x,y) \in \mathbb{R}^2$,

$$f(x,y) = \begin{cases} \text{blue} & \text{if } v(x) \geq v(y), v(x) \geq v(1) \\ \text{green} & \text{if } v(x) < v(y), v(y) \geq v(1) \\ \text{red} & \text{if } v(x) < v(1), v(y) < v(1) \end{cases}$$

This assigns each point of the plane a unique colour. The figure immediately after Theorem 5.2 shows values of this colouring at points on a 20×20 lattice overlayed across the unit square.

The next step is to prove the following lemma on the various properties of this colouring. As in the proof of theorem 5.2, we refer to any triangle with vertices coloured three different colours as a BGR triangle.

Lemma 6.2.

(1) *There is no BGR triangle in $\mathbb{R}^2$ with area equal to $\frac{1}{n}$ for any odd number n.*

(2) *Any line in $\mathbb{R}^2$ contains points of at most two colours.*

Proof. Let $p_B, p_G, p_R \in \mathbb{R}^2$ be coloured blue, green and red respectively. Recall from first year linear algebra that up to a sign

$$\text{Area of } \triangle(p_B, p_G, p_R) = \frac{1}{2} \det \begin{bmatrix} x_B & y_B & 1 \\ x_G & y_G & 1 \\ x_R & y_R & 1 \end{bmatrix}$$

Now, observe that the construction of the colouring means that the max-valued term in the expansion of this determinant lies along the diagonal. Indeed all other terms $x_i \cdot y_j \cdot 1$ must draw at least one entry from below the diagonal. Thus $v(x_B \cdot y_G \cdot 1) \geq v(x_i \cdot y_j \cdot 1)$, and recalling that v is non-Archimedean and lemma 3.14, $v(2 \cdot \text{area} \triangle(p_B, p_G, p_R) = v(x_B \cdot y_G \cdot 1) = v(x_B)v(y_G)v(1) \geq v(1)v(1)v(1) \geq 1$

Now, to prove (1), suppose the area $\triangle(p_B, p_G, p_R)$ was $\pm\frac{1}{n}$. Then we have $v(2 \cdot \text{area} \triangle(p_B, p_G, p_R) = v(\pm\frac{2}{n}) = v(\frac{1}{2})^{-1} v(\pm\frac{1}{n})$ We recall now that $v(\frac{1}{2}) > 1$ and, for odd n, $v(\pm\frac{1}{n}) = 1$. (lemma 4.3). We find therefore that $v(2 \cdot \text{area} \triangle(p_B, p_G, p_R) < 1$. This is a contradiction to it being greater or equal to 1 and we conclude that no such BGR triangles exist.

To prove (2), we simply note that if p_B, p_G, p_R lie in a line, the area of the triangle thus formed is 0. $v(0) = 0$, which also contradicts $v(2 \cdot \text{area} \triangle(p_B, p_G, p_R) \geq 1$. □

The conclusion follows swiftly from this lemma. Note that, from the definition of the colouring, $(0,0)$ is coloured red, $(0,1)$ green, $(1,0)$ blue and $(1,1)$ also blue. Together with (2) of the lemma, all the conditions for the form of Sperner's Lemma we proved in section 5 are met. We conclude that any dissection of the unit square into triangles contains at least one BGR triangle.

Using (1) of the lemma, we deduce this triangle cannot have area $\frac{1}{n}$ with n odd. But since the unit square has area 1, this means not all the triangles can have equal area. So the theorem is proved. □

.

References

[1] Martin Aigner and Gunter M. Ziegler. Proofs from The Book. Springer-Verlag, Berlin, fifth edition, 2014. Including illustrations by Karl H. Hofmann