THE BANACH-TARSKI THEOREM ON POLYGONS, AND THE CANCELLATION LAW

R. J. GARDNER AND M. LACZKOVICH

(Communicated by R. Daniel Mauldin)

ABSTRACT. The Banach–Tarski theorem on polygons in $\mathbf{R}^2$ implies that two polygons are equidecomposable if and only if they are equidissectable. The possibility of strengthening this result in various ways is investigated. We show that if two polytopes in $\mathbf{R}^d$ are equidecomposable under a finite set of isometries which generates a discrete group, then they are equidissectable using the same isometries. We then give a simple example in $\mathbf{R}$ showing that this is not true for arbitrary finite sets of isometries. A modification of this example is used to answer a question of S. Wagon concerning the cancellation law.

1. INTRODUCTION

In [1] (the paper containing the celebrated Banach–Tarski “paradox”), S. Banach and A. Tarski stated a theorem which implies the following remarkable result.

Theorem A (Banach–Tarski). *Two polygons in $\mathbf{R}^2$ are equidecomposable if and only if they are equidissectable.*

Equidecomposable means that one can be partitioned into finitely many disjoint sets, here called pieces, which can be rearranged to form a partition of the other; *equidissectability* is the classical notion where the pieces are triangles, allowed to overlap on their boundaries.

The question arises, whether this theorem can be strengthened. There are at least two ways in which this might be done. The first is to try to prove the same result for convex bodies in $\mathbf{R}^2$ instead of polygons. Of course, one has to substitute the appropriate generalization of equidissectability. Introduced in [8], this is called convex equidecomposability, and it allows the pieces themselves to be convex bodies which may overlap on their boundaries. Now, if one of the polygons is replaced by a convex body, a partial generalization of Theorem A can be proved by restricting the pieces to be moved by isometries from a

Received by the editors May 1, 1989 and, in revised forms, June 20, 1989 and October 6, 1989.
1980 *Mathematics Subject Classification* (1985 Revision). Primary 52A10; Secondary 52A20, 28C10.

Key words and phrases. Polygon, convex body, equidecomposable, equidissectable, circle-squaring, matching, cancellation law.

discrete group [2, Theorem 1]; but even this is generally false if both polygons are replaced by convex bodies (see [3]).

What Banach and Tarski actually proved is that two polygons in $\mathbf{R}^2$ are equidecomposable if and only if they have the same area; in fact, each half of the theorem follows from slightly earlier papers published by them separately. The theorem stated above follows, since the Bolyai–Gerwein theorem says that two polygons have equal areas if and only if they are equidissectable (see [10, Theorem 3.2]). It is natural to wonder whether there is a more direct proof of Theorem A. Perhaps—and this is the second way that Theorem A might be improved—the required dissection can be achieved using the same isometries that demonstrate the equidecomposability. Our first result uses Hall’s matching theorem to strengthen [2, Theorems 1 and 5], by showing that this is true when the isometries generate a discrete group (see §2 for unexplained terms).

Theorem 1. *Let F be a finite set of isometries of $\mathbf{R}^d$, which generate a discrete group G . Suppose that K_1 is a polytope and K_2 is a convex body in $\mathbf{R}^d$. The following conditions are equivalent.*

- (i) K_1 and K_2 are equidecomposable mod null sets, under isometries from F .
- (ii) K_1 and K_2 are equidecomposable mod first category sets, under isometries from F .
- (iii) K_1 and K_2 are convex equidecomposable, under isometries from F .

Theorem 1 was obtained as a contribution to Tarski’s famous problem of whether a disc and a square of equal areas are equidecomposable. Recently, in [5], the second author has shown that this is so and furthermore that only translations need be used. On the other hand, a disc and a square are clearly not convex equidecomposable. Consequently Theorem 1 (and its weaker predecessor in [2]) are generally not true if the group G is abelian, which answers the question on page 8 of [2] negatively. The following example also shows this.

Example 1. *There are two intervals in $\mathbf{R}$ which are equidecomposable under a certain finite set F of isometries, but which are not equidissectable under isometries from the group generated by F .*

The point is that the methods of [5] require a huge number of translations, and the proofs are very complicated. By contrast, Example 1 uses only four isometries, and is based on the relatively simple technique of [4].

Example 1 also provides a simple way to see that the second possible strong form of Theorem A suggested above does not exist. For, by taking the cartesian product of the two intervals in Example 1 with the unit interval, one obtains two squares which are equidecomposable under isometries in F , but not equidissectable by those isometries, or even isometries from the group generated by F . (Again, the much more difficult methods of [5] show more; in fact, any two polygons of equal areas are equidecomposable by translations.)

The ideas behind Example 1 may also be used to answer a question [10, p. 231, Question 14] of S. Wagon concerning the cancellation law.

Suppose G is a group of isometries of $\mathbf{R}^d$ and $\mathcal{A}$ is a G -invariant algebra of subsets of $\mathbf{R}^d$. The cancellation law in its simplest case demands that whenever A_1, A_2 are disjoint and G -equidecomposable sets in $\mathcal{A}$, similarly B_1, B_2 are disjoint and G -equidecomposable sets in $\mathcal{A}$, and the sets $A_1 \cup A_2$ and $B_1 \cup B_2$ are also G -equidecomposable in $\mathcal{A}$, then so are A_1 and B_1 . More generally, the pairs of sets considered above may be replaced by n -tuples. In a more abstract setting the cancellation law can be expressed algebraically using the type semigroup introduced by A. Tarski (for the definitions see [10, Chapter 8]).

It is known that the general cancellation law holds when $\mathcal{A}$ is the power set [10, Theorem 8.7]. However, for certain proper subalgebras (such as the algebra of Borel sets), it is unknown if the cancellation law holds. Our example provides an algebra for which it does not.

Wagon's question actually arises from a still more general context in which isometries are replaced by arbitrary bijections. In this setting the cancellation law has been shown to fail by J. K. Truss in [9]. Although his example is stronger in that it applies to a σ -algebra, ours uses isometries in $\mathbf{R}$. Further examples of the latter type are given in [6].

Example 2. *There is a group G acting on a subset X of $\mathbf{R}$ and a G -invariant subalgebra $\mathcal{A}$ of the power set $\mathcal{P}(X)$ such that the cancellation law fails for G -equidecomposability with pieces in $\mathcal{A}$.*

2. DEFINITIONS AND NOTATION

We denote the interior, boundary, and cardinality of a set E by $\text{int } E$, $\text{bd } E$, and $|E|$, respectively.

Two sets A, B in $\mathbf{R}^d$ are *equidecomposable* if there are disjoint decompositions $A = \bigcup_{i=1}^k A_i$, $B = \bigcup_{i=1}^k B_i$, and isometries τ_i of $\mathbf{R}^d$ such that $B_i = \tau_i A_i$ for $i = 1, \dots, k$. If G is a group and $\tau_i \in G$ for $i = 1, \dots, k$, we say that A and B are *G -equidecomposable*.

We denote Lebesgue measure in $\mathbf{R}^d$ by λ_d . If, instead of disjointness, the sets above only satisfy $\lambda_d(A_i \cap A_j) = \lambda_d(B_i \cap B_j) = 0$ (respectively, $A_i \cap A_j$ and $B_i \cap B_j$ are of first category), for $1 \leq i \neq j \leq k$, we say A and B are *equidecomposable mod null sets (mod first category sets)*.

If A and B are polytopes in $\mathbf{R}^d$ and the sets A_i and B_i are simplices satisfying, instead of disjointness, $\text{int } A_i \cap \text{int } A_j = \text{int } B_i \cap \text{int } B_j = \emptyset$, $1 \leq i \neq j \leq k$, then A and B are *equidissectable*. Finally, convex bodies A and B are *convex equidecomposable* if the latter condition on the sets A_i and B_i holds, where these are also convex bodies.

A group G of isometries is *discrete* if, for every bounded subset C of $\mathbf{R}^d$, $C \cap g(C) = \emptyset$ except for a finite number of g in G .

3. PROOFS OF THE RESULTS

Proof of Theorem 1. Since (iii) $\Rightarrow$ (i) or (ii) is obvious, we shall assume (i) or (ii). In either case, we begin by following the argument of [2, Theorem 1]. Here it is shown that K_2 must be a polytope, and there is a dissection of K_1 into subpolytopes P_{jk} , $1 \leq j \leq l$, $1 \leq k \leq p_j$, such that:

- (a) For each j , all the polytopes P_{jk} are congruent, for $1 \leq k \leq p_j$.
- (b) For each j, k , there is an isometry $g_{jk} \in G$ such that the polytopes $Q_{jk} = g_{jk}P_{jk}$, $1 \leq j \leq l$, $1 \leq k \leq p_j$, form a dissection of K_2 .
- (c) For each j , $1 \leq j \leq l$, $\bigcup\{P_{jk}: 1 \leq k \leq p_j\}$ and $\bigcup\{Q_{jk}: 1 \leq k \leq p_j\}$ are equidecomposable (mod null sets or first category sets, as appropriate) under isometries from F .

Our assumptions imply that there is a set N such that $(K_1 - N)$ and $(K_2 - N)$ are equidecomposable under isometries from F , where $\lambda_d(N) = 0$ or N is of first category, according to whether (i) or (ii) holds. For each $x \in (K_1 - N)$, there is a $\tau_x \in F$ such that $\tau_x(x) \in (K_2 - N)$ is the point corresponding to x under the decomposition. Define ψ on $(K_1 - N)$ by letting $\psi(x) = \tau_x(x)$ for each x .

Let us fix a j , $1 \leq j \leq l$. We shall construct a bipartite graph Γ as follows. The nodes of the two parts of Γ are the polytopes P_{jk} , $1 \leq k \leq p_j$, and Q_{jk} , $1 \leq k \leq p_j$, respectively. Two nodes P_{jk_1} and Q_{jk_2} are joined by an edge if there is an $x \in P_{jk_1}$ such that $\psi(x) \in Q_{jk_2}$.

As in [2, Theorem 5], we denote by $\bar{\mu}$ a finitely additive, G -invariant measure defined on all subsets of $\mathbf{R}^d$. If (i) is assumed, we take $\bar{\mu}$ to be an extension of λ_d ; if (ii) holds, we suppose that $\bar{\mu}$ satisfies $\bar{\mu}(E) = \lambda_d(E)$ if $\lambda_d(\text{bd } E) = 0$, and $\bar{\mu}(H) = 0$ for all first category sets H .

Let I be a subset of $\{1, \dots, p_j\}$, and let $X = \bigcup\{P_{jk}: k \in I\} - N$. Then, by (a),

$$\bar{\mu}(X) = |I| \cdot \lambda_d(P_{jk}),$$

since $\bar{\mu}(N) = 0$. Now the set $\psi(X) - N$ is equidecomposable with X , under isometries in F , so must have the same $\bar{\mu}$ -measure. Since $\bar{\mu}(P_{jk}) = \lambda_d(P_{jk})$ for all k , it follows by (b) and (c) that $[\psi(X) - N] \cap Q_{jk} \neq \emptyset$ for all k belonging to a subset J of $\{1, \dots, p_j\}$ satisfying $|J| \geq |I|$. But $k' \in J$ if and only if $Q_{jk'}$ is joined by an edge in Γ to some P_{jk} with $k \in I$. Consequently Γ satisfies the hypothesis of Hall's matching theorem (see, for example, [7, Theorem 1.1.3]), and Γ has a matching.

The existence of this matching shows immediately that the sets

$$\bigcup\{P_{jk}: 1 \leq k \leq p_j\} \quad \text{and} \quad \bigcup\{Q_{jk}: 1 \leq k \leq p_j\}$$

are equidissectable under isometries from F . Repeating the argument for each j , $1 \leq j \leq l$, we conclude that (iii) is true. $\square$

Construction of Example 1. Let $u \in [0, 1]$ be irrational. From a matching constructed in [4] it follows immediately that $[0, 1]$ is equidecomposable with

itself under the isometries $\sigma_1(x) = x + u$, $\sigma_2(x) = x - u$, $\sigma_3(x) = -x + u$, and $\sigma_4(x) = -x - u + 2$. Suppose that v is a real number such that 1, u , and v are rationally independent. Then the intervals $[0, 1]$ and $[v, v + 1]$ are equidecomposable under the isometries $\tau_i(x) = \sigma_i(x) + v$, $1 \leq i \leq 4$.

Let G be the group of all isometries of the form $g(x) = \pm x + ku + lv + 2m$, where k , l , and m are integers with $(k + l)$ even. We shall prove that $[0, 1]$ and $[v, v + 1]$ are not equidissectable under isometries from G .

Suppose the contrary. Then there are partitions $0 = a_0 < a_1 < \dots < a_n = 1$ of $[0, 1]$ and $v = b_0 < b_1 < \dots < b_n = v + 1$ of $[v, v + 1]$, such that the intervals $[a_i, a_{i+1}]$, $0 \leq i < n$, are in one-to-one correspondence with the intervals $[b_j, b_{j+1}]$, $0 \leq j < n$, via isometries from G . Denote by g_i the isometry in G which maps $[a_i, a_{i+1}]$ to some $[b_j, b_{j+1}]$, for $0 \leq i < n$.

Form a bipartite graph Δ as follows. The nodes of each part of Δ are the numbers a_i , $0 \leq i \leq n$, and b_j , $0 \leq j \leq n$, respectively. We join a_i and b_j by an edge if $g_i(a_i) = b_j$ and $i < n$, or $g_{i-1}(a_i) = b_j$ and $i > 0$. Then a_0, a_n, b_0 , and b_n have degree one, and all other points have degree two (we allow multiple edges). Therefore a_0 belongs to a component Δ_0 of Δ such that each point in Δ_0 has degree two, unless it is a_0, a_n, b_0 , or b_n . So there is a path in Δ_0 beginning at a_0 and ending at a_n, b_0 , or b_n .

Since $a_0 = 0$, all points in this path correspond to numbers of the form $(ku + lv + 2m)$ with $(k + l)$ even. However, since 1, u , and v are rationally independent, none of the numbers $a_n = 1$, $b_0 = v$, or $b_n = v + 1$ can be of this form. This contradiction proves our assertion above.

Construction of Example 2. Suppose the real numbers u , v and the group G are as in the construction of Example 1. Let $X = \mathbf{R} - H$, where $H = \{ku + lv + m : k, l, m \in \mathbf{Z}\}$, and let $\mathcal{A}$ be the algebra generated by the sets $(a, b) \cap X$, $a, b \in H$. We first claim that the sets $B = [(0, 1) \cap X] \cup [(2, 3) \cap X]$ and $C = [(v, v + 1) \cap X] \cup [(v + 2, v + 3) \cap X]$ are G -equidecomposable with pieces in $\mathcal{A}$.

To see this, note that $B = \bigcup_{i=1}^4 B_i$ and $C = \bigcup_{i=1}^4 C_i$, where $B_1 = (0, u) \cap X$, $B_2 = (u, 1) \cap X$, $B_3 = (2, 3 - u) \cap X$, $B_4 = (3 - u, 3) \cap X$, and $C_1 = (v, v + u) \cap X$, $C_2 = (v + 2, v + 3 - u) \cap X$, $C_3 = (v + u, v + 1) \cap X$, $C_4 = (v + 3 - u, v + 3) \cap X$. Now $C_i = g_i B_i$, $1 \leq i \leq 4$, where $g_1(x) = -x + u + v$, $g_2(x) = x - u + v + 2$, $g_3(x) = x + u + v - 2$, and $g_4(x) = -x - u + v + 6$.

It is clear that $(0, 1) \cap X$ and $(2, 3) \cap X$ are G -equidecomposable in $\mathcal{A}$, and also that $(v, v + 1) \cap X$ and $(v + 2, v + 3) \cap X$ are G -equidecomposable in $\mathcal{A}$; the single translation $g(x) = x + 2$ serves in each case. On the other hand, $(0, 1) \cap X$ and $(v, v + 1) \cap X$ are not G -equidecomposable in $\mathcal{A}$, since, if they were, the intervals $[0, 1]$ and $[v, v + 1]$ would be equidissectable under isometries from G . As we saw in Example 1, this is not the case.

REFERENCES

1. S. Banach and A. Tarski, *Sur la décomposition des ensembles des points en parties respectivement congruentes*, Fund. Math. **6** (1924), 244–277.
2. R. J. Gardner, *Convex bodies equidecomposable by locally discrete groups of isometries*, Mathematika **32** (1985), 1–9.
3. —, *A problem of Sallee on equidecomposable convex bodies*, Proc. Amer. Math. Soc. **94** (1985), 329–332.
4. M. Laczkovich, *Closed sets without measurable matching*, Proc. Amer. Math. Soc. **103** (1988), 894–896.
5. —, *Equidecomposability and discrepancy; a solution of Tarski's circle-squaring problem*, J. Reine Angew. Math. **404** (1990), 77–117.
6. —, *Invariant signed measures and the cancellation law*, preprint.
7. L. Lovász and M. D. Plummer, *Matching theory*, North-Holland, Amsterdam, 1986.
8. G. T. Sallee, *Are equidecomposable plane convex sets convex equidecomposable?*, Amer. Math. Monthly **76** (1969), 926–927.
9. J. K. Truss, *The failure of cancellation laws for equidecomposability types*, Canad. J. Math. (to appear).
10. S. Wagon, *The Banach–Tarski paradox*, Cambridge University Press, New York, 1985.

ISTITUTO ANALISI GLOBALE E APPLICAZIONI, VIA S. MARTA 13/A, 50139 FIRENZE, ITALY

Current address: Mathematisches Institut II, Universität Karlsruhe, Englerstrasse 2, 7500 Karlsruhe, West Germany

DEPARTMENT OF ANALYSIS, EÖTVÖS LORÁND UNIVERSITY, 1088 BUDAPEST, MUZEUM KRT. 6–8, HUNGARY