Discrete Comput Geom (2010) 43: 393–401
DOI 10.1007/s00454-009-9179-x

On a Question of Erdős and Ulam

Jozsef Solymosi · Frank de Zeeuw

Received: 18 June 2008 / Revised: 10 April 2009 / Accepted: 11 April 2009 /
Published online: 7 May 2009

Abstract Ulam asked in 1945 if there is an everywhere dense *rational set*, i.e., 1 a point set in the plane with all its pairwise distances rational. Erdős conjectured that if a set S has a dense rational subset, then S should be very special. The only known types of examples of sets with dense (or even just infinite) rational subsets are lines and circles. In this paper we prove Erdős' conjecture for algebraic curves by showing that no irreducible algebraic curve other than a line or a circle contains an infinite rational set.

Keywords Rational distances · Erdős problems in discrete geometry · Rational points

1 Introduction

We define a *rational set* to be a set $S \subset \mathbb{R}^2$ such that the distance between any two elements is a rational number. We are interested in the existence of infinite rational distance sets on algebraic curves.

On any line, one can easily find an infinite rational set that is in fact dense. It is also an easy exercise to find an everywhere dense rational subset of the unit circle. On the other hand, it is not known if there is a rational set with 8 points *in general position*, i.e., no 3 on a line, no 4 on a circle. In 1945, Anning and Erdős [1] proved that any infinite *integral* set, i.e., where all distances are integers, must be contained in a line. Problems related to rational and integral sets became one of Erdős' favorite subjects in combinatorial geometry [6–9, 11, 12].

The first author was supported by NSERC and OTKA grants and by Sloan Research Fellowship.

J. Solymosi (✉) · F. de Zeeuw
Department of Mathematics, UBC, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada
e-mail: solymosi@math.ubc.ca

F. de Zeeuw
e-mail: fdezeeuw@math.ubc.ca

In 1945, when Ulam heard Erdős' simple proof [5] of his theorem with Anning, he said that he believed there is no everywhere dense rational set in the plane, see Problem III.5 in [22] and also [10]. Erdős conjectured that an infinite rational set must be very restricted, but that it was probably a very deep problem [10, 11]. Not much progress has been made on Ulam's question. There were attempts to find rational sets on parabolas [3, 4], and there were some results on integral sets, in particular bounds were found on the diameter of integral sets [15, 21]. Very recently Kreisel and Kurz [18] found an integral set with 7 points in general position.

In this paper, we prove that lines and circles are the only irreducible algebraic curves that contain infinite rational sets. Our main tool is Faltings' Theorem [13]. We will also show that if a rational set S has infinitely many points on a line or on a circle, then all but 4 resp. 3 points of S are on the line or on the circle. This answers questions of Guy, Problem D20 in [14], and Pach, Sect. 5.11 in [2].

2 Main Result

Our main result is the following.

Theorem 2.1 *Every rational set of the plane has only finitely many points in common with an algebraic curve defined over* $\mathbb{R}$, *unless the curve has a component which is a line or a circle.*

The two special cases, line and circle, are treated in more detail in the next theorem.

Theorem 2.2 *If a rational set S has infinitely many points on a line or on a circle, then all but* 4 *resp.* 3 *points of S are on the line or on the circle.*

Note that there are infinite rational sets with all but 4 points on a line, and there are infinite rational sets with all but 3 points on a circle. The circle case follows from the line case by applying an inversion with rational radius and center one of the 4 points not on the line. A construction of Huff [16, 19] gives an infinite rational set with all but 4 points on a line.

We can formulate our Theorem 2.1 in a different way by using the term *curve-general position*: we say that a point set S of $\mathbb{R}^2$ is in curve-general position if no algebraic curve of degree d contains more than $d(d+3)/2$ points of S. Note that $d(d+3)/2$ is the number of points in general position that determine a unique curve of degree d.

Corollary 2.3 *If S is an infinite rational set in general position, then there is an infinite $S' \subset S$ such that S' is in curve-general position.*

Proof Let S_5 consist of any five points in S, and let T_5 be the set of finitely many points on the unique conic through those five points. Continue recursively; at step n, add a point from $S \backslash T_{n-1}$ to S_{n-1} to get S_n. For each d such that $d(d+3)/2 \leq n$, let T_n be the union of T_{n-1} and the set of points of S that are on a curve of degree

d through any $d(d+3)/2$ points in S_n. Since each T_n is finite, we can always add another point. Then the infinite union of the sets S_n is an infinite subset of S with the required property. □

3 Proof of Theorem 2.1

3.1 General Approach

We will use the following theorem of Faltings [13].

Theorem (Faltings) *A curve of genus* ≥ 2, *defined over a number field*, *contains only finitely many rational points*.

In this paper by *curve* (defined over a field $K \subset \mathbb{R}$) we usually mean the zero set in $\mathbb{R}^2$ of a polynomial in two variables with coefficients from K. However, when we consider the genus of a curve, we are actually talking about the projective variety defined by the polynomial. For definitions, see [20].

First suppose that we have an infinite rational set S contained in a curve C of genus ≥ 2, defined over $\mathbb{R}$. We can move two points in S to $(0,0)$ and $(0,1)$, so that by Lemma 3.2 below the elements of S are of the form $(r_1, r_2\sqrt{k})$. Then by the remark after Lemma 3.2, the curve is defined over $\mathbb{Q}(\sqrt{k})$. By Faltings' theorem, S must be finite.

Below we will show that if we have an infinite rational set S on a curve C_1 of genus 0 or 1, then all but finitely many of the points in S will in fact give points on a curve C_2 in $\mathbb{R}^3$ of genus ≥ 2. More precisely, assuming that $(0,0)$ and $(0,1)$ are in S, a point $(r_1, r_2\sqrt{k})$ will give a point $(r_1, r_2\sqrt{k}, r_3)$ on a curve C_2, with all the r_i rational. Again we conclude by Faltings' theorem that the original set S could not have been infinite.

3.2 Two Lemmata

Rationality of distances in $\mathbb{R}^2$ is clearly preserved by translations, rotations, and uniform scaling ($(x,y) \mapsto (\lambda x, \lambda y)$ with $\lambda \in \mathbb{Q}$). More surprisingly, rational sets are preserved under certain central inversions. This will be an important tool in our proof below.

Lemma 3.1 *If we apply inversion to a rational set* S, *with center a point* $x \in S$ *and rational radius*, *then the image of* $S\backslash\{x\}$ *is a rational set*.

Proof We may assume the center to be the origin and the radius to be 1. The properties of inversion are most easily seen in complex notation, where the map is $z \mapsto 1/z$. Suppose that we have two points z_1, z_2 with rational distances $|z_1|$, $|z_2|$ from the origin and with $|z_2 - z_1|$ rational. Then

$$\left|\frac{1}{z_1} - \frac{1}{z_2}\right| = \left|\frac{z_2 - z_1}{z_1 z_2}\right| = \frac{|z_2 - z_1|}{|z_1||z_2|}$$

is also rational. □

A priori, points in a rational set could take any form. However, after moving two of the points to two fixed rational points by translating, rotating, and scaling, the points are almost rational points. The following simple lemma is well known among researchers working with integer sets. As far as we know, it was proved first by Kemnitz [17].

Lemma 3.2 *For any rational set S, there is a square-free integer k such that if a similarity transformation T transforms two points of S into $(0,0)$ and $(1,0)$, then any point in $T(S)$ is of the form*

$$\left(r_1, r_2\sqrt{k}\right), \quad r_1, r_2 \in \mathbb{Q}.$$

Note that this implies that any curve of degree d containing at least $d(d+3)/2$ points from $T(S)$ is defined over $\mathbb{Q}(\sqrt{k})$.

3.3 Curves of Genus 1

Let $C_1 : f(x,y) = 0$ be an irreducible algebraic curve of genus $g_1 = 1$ and degree $d \geq 3$. Suppose that there is an infinite set S on C_1 with pairwise rational distances. Assume that the points $O = (0,0)$ and $(1,0)$ are on C_1 and in S and that O is not a singularity of C_1. Below we will be allowed to make any other assumptions on C_1 that we can achieve by translating, rotating, or scaling it, as long as we also satisfy the assumptions above. In particular, we can use any of the infinitely many rotations about the origin that put a different point of S on the x-axis.

We wish to show that the intersection curve C_2 of the surfaces

$$X : f(x,y) = 0,$$
$$Y : x^2 + y^2 = z^2,$$

has genus $g_2 \geq 2$.

Consider C_1 as a curve in the $z = 0$ plane, and define the map $\pi : C_2 \to C_1$ by $(x,y,z) \mapsto (x,y)$, i.e., the restriction to C_2 of the vertical projection from the cone Y to the $z = 0$ plane. The preimage of a point (x,y) consists of the two points $(x, y, \pm\sqrt{x^2+y^2})$, except when $x^2 + y^2 = 0$, which in $\mathbb{C}^2$ happens on the two lines $x + iy = 0$ and $x - iy = 0$. Then we can determine (or at least bound from below) the genus of C_2 using the Riemann–Hurwitz formula [20] applied to π,

$$2g_2 - 2 \geq \deg \pi \cdot (2g_1 - 2) + \sum_{P \in C_2} (e_P - 1).$$

This is usually stated with equality for smooth curves, but we are allowing C_1 and C_2 to have singularities. To justify our use of it, observe that the map π corresponds to a map $\tilde{\pi} : \tilde{C}_1 \to \tilde{C}_2$ between the normalizations of the curves, for which Riemann–Hurwitz holds. The normalizations have the same genera as the original curves, and

$\tilde{\pi}$ has the same degree. Furthermore a ramification point of π away from any singularities gives a ramification point of $\tilde{\pi}$. It is enough for our purposes to have this inequality, but there could be more ramification points for $\tilde{\pi}$, above where the singularities used to be.

Applying this formula with $g_1 = 1$, $d = 2$, we have

$$g_2 \geq 1 + \frac{1}{2} \sum_{P \in C_2} (e_P - 1),$$

so to get $g_2 \geq 2$, we only need to show that π has some ramification point.

The potential ramification points are above the intersection points of C_1 with the lines $x \pm iy = 0$, of which there are $2d$ by Bézout's theorem, counting with multiplicities. Such an intersection point P can only fail to have a ramification point above it if the curve has a singularity at P or if the curve is tangent to the line there. We will show that there are only finitely many lines through the origin on which one of those two things happens. Then certainly one of the infinitely many rotations of C_1 that we allowed above will give an intersection point of C_1 with $x \pm iy = 0$ that has a ramification point above it.

The intersection of a line $y = ax$ with $f(x, y) = 0$ is given by $p_a(x) = f(x, ax) = 0$, and if the point of intersection is a singularity or a point of tangency, then $p_a(x)$ has a multiple root. We can detect such multiple roots by taking the discriminant of $p_a(x)$, which will be a polynomial in a that vanishes if and only if $p_a(x)$ has a multiple root. Hence for all but finitely many values of a, the line $y = ax$ has d simple intersection points with $f(x, y) = 0$. So indeed there is an allowed rotation after which π is certain to have a ramification point.

3.4 Curves of Genus 0, $d \geq 4$

Let $C_1 : f(x, y) = 0$ be an irreducible algebraic curve of genus $g_1 = 0$, and again assume that it passes through the origin but does not have a singularity there. Then Riemann–Hurwitz with the same map π as above gives

$$g_2 \geq -1 + \frac{1}{2} \sum_{P \in C_2} (e_P - 1),$$

so to get $g_2 \geq 2$ we need to show that there are at least 5 ramification points. As above, we can ensure that the lines $x \pm iy$ each have d simple points of intersection. Discounting the intersection point of the two lines, this gives $2d - 2$ ramification points. Hence if the degree of f is $d \geq 4$, we are done.

3.5 Curves of Genus 0, $d = 2, 3$

Let $d = 3$ and assume that $f(x, y) = 0$ is not a line or a circle. Consider applying inversion with the origin as center to the curve. This is a birational transformation, so does not change the genus. Therefore, when inversion increases the degree of f to above 4, we are done.

Algebraically, inversion in the circle around the origin with radius 1 is given by

$$(x, y) \mapsto \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right),$$

and since this map is its own inverse, the curve $f(x, y) = 0$ is sent to the curve

$$C_3 : \left(x^2+y^2\right)^k \cdot f\left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right) = 0,$$

where $k \le d$ is the lowest integer that makes this a polynomial. This curve is irreducible if and only if the original curve is irreducible. Since f does not have a singularity at the origin, it has a linear term $ax + by$ with a, b not both zero. After inversion this gives a highest-degree term

$$(ax+by)\left(x^2+y^2\right)^{k-1}.$$

In our situation, $d = 3$, so if $k = 3$, the curve C_3 has degree $2k - 1 = 5$, and we are done.

The only other possibility is that $k = 2$, which happens if $x^2 + y^2$ divides the leading terms of f. We will treat these cases in a completely different way.

If $d = 2$, then applying inversion will give a curve of degree 3, unless its leading terms are $x^2 + y^2$, which exactly means that it is a circle! So we treat this case by reducing it to the $d = 3$ case.

Since f has degree 3 and genus 0, it must have a singularity. The singularity need not be in our rational set, but it is always a rational point, so we can move it to the origin, while maintaining the almost-rational form of the points in our rational set. Then f must have the form

$$(ax+by)\left(x^2+y^2\right) + cx^2 + dy^2 + exy.$$

Note that this is exactly what we get if we apply inversion to a quadratic that is not a circle and goes through the origin.

In fact, we can ensure that $(1, 0)$ is on the curve again, so that $a + c = 0$. Then if we divide by c, f is of the form

$$(-x+by)\left(x^2+y^2\right) + x^2 + dy^2 + exy.$$

We can parameterize this curve using lines $x = ty$, giving the parameterization

$$y(t) = \frac{t^2+et+d}{(t-b)(t^2+1)} =: \frac{p(t)}{q(t)}, \qquad x(t) = t \cdot y(t).$$

If we let t_j be a value of t that gives one of the points from our rational distance set, it follows that for infinitely many t,

$$\left(y(t)-y(t_j)\right)^2 + \left(x(t)-y(t_j)\right)^2 = \left(\frac{p(t)}{q(t)} - \frac{p(t_j)}{q(t_j)}\right)^2 + \left(t \cdot \frac{p(t)}{q(t)} - t_j \cdot \frac{p(t_j)}{q(t_j)}\right)^2$$

is a square. Then we can multiply by $q(t)^2 q(t_j)^2$ to get infinitely many squares of the form

$$\big(p(t)q(t_j) - p(t_j)q(t)\big)^2 + \big(tp(t)q(t_j) - t_j p(t_j)q(t)\big)^2.$$

This polynomial has degree 6 in t. It has a factor $(t - t_j)^2$ and a factor $t^2 + 1$, since taking $t = \pm i$ gives (using $q(\pm i) = 0$)

$$\big(p(\pm i)q(t_j)\big)^2 + \big(\pm i \cdot p(\pm i)q(t_j)\big)^2 = 0.$$

Factoring these out, we get a quadratic polynomial $Q_j(t)$ in t. Its leading coefficient is

$$\big(t_j^2 + 1\big)\big(\big(d^2 + b^2\big)t_j^2 + 2\big(b^2 e + db - d^2 b\big)t_j + b^2 e^2 + b^2 d^2 + d^2 + 2ebd\big),$$

and its constant term is

$$\big(t_j^2 + 1\big)\big(\big(1 + (e + b)^2\big)t_j^2 + 2(bd - b + de)t_j + d^2 + b^2\big).$$

These polynomials in t_j are not identically zero (if b and d were both 0, then f would be reducible), hence we can pick t_j so that they are not zero. Then in turn $Q_j(t)$ is a proper quadratic polynomial, and since it is essentially a distance function in the real plane, it cannot have real roots, so it has two distinct imaginary roots.

Therefore our infinite rational set gives infinitely many solutions to the equations

$$z_j^2 = \big(t^2 + 1\big) \cdot Q_j(t).$$

Multiplying three of these together, and moving $(t^2 + 1)^2$ into the square on the left, we get infinitely many solutions to

$$z^2 = \big(t^2 + 1\big) Q_1(t) Q_2(t) Q_3(t).$$

If there are no multiple roots on the right, then this is a hyperelliptic curve of degree 8, so it has genus 3, hence cannot have infinitely many solutions, a contradiction.

The one thing we need to check is that we can choose the t_j so that the Q_j do not have roots in common. We need some notation: write

$$Q_j(t) = c_2(t_j)t^2 + c_1(t_j)t + c_0(t_j),$$

where

$$\begin{aligned}
c_2(t_j) &= \big(1 + (e + b)^2\big)t_j^2 + 2(bd + de - b)t_j + d^2 + b^2 \\
c_1(t_j) &= 2(bd + de - b)t_j^2 + 2\big(b^2 + d^2 - bed - bd - be - d\big)t_j \\
&\quad + 2\big(bd + b^2 e - bd^2\big) \\
c_0(t_j) &= \big(d^2 + b^2\big)t_j^2 + 2\big(b^2 e + db - d^2 b\big)t_j + b^2 e^2 + b^2 d^2 + d^2 + 2ebd.
\end{aligned}$$

Suppose that for infinitely many t_j, the polynomial $Q_j(t)$ has the same roots x_1 and x_2. Then for each of those t_j, we have

$$c_1(t_j) = -(x_1 + x_2) \cdot c_2(t_j), \qquad c_0(t_j) = x_1 \cdot x_2 \cdot c_2(t_j).$$

If we look at the coefficients of the t_j terms in these equations, we see that

$$-x_1 - x_2 = \frac{2(b^2 + d^2 - bed - bd - be - d)}{2(bd - b + de)} = -b - \frac{be + d - d^2}{bd + de - b},$$
$$x_1 \cdot x_2 = \frac{2(b^2 e + db - d^2 b)}{2(bd + de - b)} = b \cdot \frac{be + d - d^2}{bd + de - b}.$$

Here we can read off that the roots are $x_1 = b$ and $x_2 = \frac{be+d-d^2}{bd+de-b}$, which is a contradiction, since the roots had to be imaginary.

4 Proof of Theorem 2.2

We will prove that if a rational set has infinitely many points on a line, then it can have at most 4 points off the line. The corresponding statement for 3 points off a circle then follows by applying an inversion. More precisely, suppose that we have a rational set S with infinitely many points on a circle C and at least 4 points off that circle. Assume that the origin is one of the points in $S \cap C$, and apply inversion with the origin as center and with some rational radius. That turns C into a line L, and we get a rational set with infinitely many points on L and 4 other points. Moreover, the new origin can be added to S, so that we get 5 points off the line, contradicting what we will prove below. To see that the new origin has rational distance to all points in S, observe that in complex notation the distances $|z|$ to the old origin were rational for all $z \in S$ and that the distances to the new origin are $1/|z|$.

To prove the statement for a line, our main tool will again be Faltings' theorem, but now applied to the hyperelliptic curve

$$y^2 = \prod_{i=1}^{6} (x - \alpha_i),$$

which has genus 2 if and only if the α_i are distinct.

Suppose that we have a rational set S with infinitely many points on a line, say the x-axis, and 5 or more points off that line. Then we can assume that 3 of those points are above the x-axis and that one of them is at $(0, 1)$. Let the other two points be at (a_1, b_1) and (a_2, b_2). Note that we are taking 3 points on one side of the line, because we want to avoid having one point a reflection of another. If we had, say, $(a_1, b_1) = (0, -1)$, the argument below would break down.

Take a point $(x, 0)$ of S on the x-axis with $x \neq 0, a_1, a_2$. Then we have that

$$x^2 + 1, \qquad (x - a_1)^2 + b_1^2, \quad \text{and} \quad (x - a_2)^2 + b_2^2$$

are rational squares, so that we get a rational point (x, y) on the curve

$$y^2 = \big(x^2+1\big)\big((x-a_1)^2+b_1^2\big)\big((x-a_2)^2+b_2^2\big).$$

This is a curve of genus 2, since the roots on the right-hand side are distinct: they are $\pm i$ and $x = a_i \pm \sqrt{-b_i^2}$ for $i = 1, 2$, which are distinct by the assumptions on the points (a_i, b_i).

Therefore the curve has genus 2 and cannot contain infinitely many rational points, contradicting the fact that S has infinitely many points on the line.

Acknowledgements We thank Kalle Karu for the useful discussions. We are also indebted to an anonymous referee who noticed that the $d = 3$ case in Sect. 3.7 was not completely covered in the previous version of the paper.

References

1. Anning, N.H., Erdős, P.: Integral distances. Bull. Am. Math. Soc. **51**, 598–600 (1945)
2. Brass, P., Moser, W., Pach, J.: Research Problems in Discrete Geometry, 1st edn. Springer, Berlin (2005). XII, 499 p.
3. Campbell, G.: Points on $y = x^2$ at rational distance. Math. Comput. **73**, 2093–2108 (2004)
4. Choudhry, A.: Points at rational distances on a parabola. Rocky Mt. J. Math. **36**(2), 413–424 (2006)
5. Erdős, P.: Integral distances. Bull. Am. Math. Soc. **51**, 996 (1945)
6. Erdős, P.: Verchu niakoy geometritchesky zadatchy. Fiz.-Mat. Spis. Bŭlgar. Akad. Nauk **5**(38), 205–212 (1962). (On some geometric problems, in Bulgarian)
7. Erdős, P.: On some problems of elementary and combinatorial geometry. Ann. Mat. Pura Appl. (IV) **CIII**, 99–108 (1975)
8. Erdős, P.: Néhány elemi geometriai problémáról. Középisk Mat. Lapok **61**, 49–54 (1980). (On some problems in elementary geometry, in Hungarian)
9. Erdős, P.: Combinatorial problems in geometry. Math. Chron. **12**, 35–54 (1983)
10. Erdős, P.: Ulam, the man and the mathematician. J. Graph Theory **9**(4), 445–449 (1985) Also appears in Creation Math. 19 (1986), 13–16
11. Erdős, P.: Some combinatorial and metric problems in geometry. In: Colloquia Mathematica Societatis János Bolyai, vol. 48, pp. 167–177. Intuitive Geometry, Siófok (1985)
12. Erdős, P., Purdy, G.B.: Extremal problems in combinatorial geometry. In: Graham, R.L., Grötschel, M., Lovász, L. (eds.) Handbook of Combinatorics, pp. 809–875. Elsevier, Amsterdam (1995)
13. Faltings, G.: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math. **73**(3), 349–366 (1983). (Finiteness theorems for abelian varieties over number fields)
14. Guy, R.: Unsolved Problems in Number Theory, 3rd edn. Problem Books in Mathematics Subseries: Unsolved Problems in Intuitive Mathematics, vol. 1. Springer, Berlin (2004). XVIII, 438 p.
15. Harborth, H., Kemnitz, A., Möller, M.: An upper bound for the minimum diameter of integral point sets. Discrete Comput. Geom. **9**(4), 427–432 (1993)
16. Huff, G.B.: Diophantine problems in geometry and elliptic ternary forms. Duke Math. J. **15**, 443–453 (1948)
17. Kemnitz, A.: Punktmengen mit ganzzahligen Abständen. Habilitationsschrift, TU Braunschweig (1988)
18. Kreisel, T., Kurz, S.: There are integral heptagons, no three points on a line, no four on a circle. Discrete Comput. Geom. **39**(4), 786–790 (2008)
19. Peeples, W.D. Jr.: Elliptic curves and rational distance sets. Proc. Am. Math. Soc. **5**, 29–33 (1954)
20. Silverman, J.: The Arithmetic of Elliptic Curves. Springer, Berlin (1986)
21. Solymosi, J.: Note on integral distances. Discrete Comput. Geom. **30**(2), 337–342 (2003)
22. Ulam, S.M.: A Collection of Mathematical Problems. Interscience Tracts in Pure and Applied Mathematics, vol. 8. Interscience, New York (1960). XIII, 150 p.