

# Polynomials vanishing on grids: The Elekes-Rónyai problem revisited\*

Orit E. Raz<sup>†</sup>      Micha Sharir<sup>‡</sup>      József Solymosi<sup>§</sup>

January 29, 2014

## Abstract

In 2000, Elekes and Rónyai [8] considered the following problem. Let  $A, B$  be two sets, each of  $n$  real numbers, and let  $f$  be a real bivariate polynomial of some constant degree. They showed that if  $|f(A \times B)| \leq cn$ , for some constant  $c$  that depends on  $\deg f$ , and for  $n \geq n_0(c)$ , for sufficiently large threshold  $n_0(c)$  that depends on  $c$ , then  $f$  must be of one of the special forms  $f(u, v) = h(\varphi(u) + \psi(v))$ , or  $f(u, v) = h(\varphi(u) \cdot \psi(v))$ , for some univariate polynomials  $\varphi, \psi, h$  over  $\mathbb{R}$ . A refined analysis and generalizations of this result are given by Elekes and Szabó [10].

In a variant of this setup, we have, in addition to  $A, B$  and  $f$ , another set  $C$  of  $n$  real numbers, and we denote by  $M$  the number of triples  $(a, b, c) \in A \times B \times C$  such that  $c = f(a, b)$ . We show that either  $M = O(n^{11/6})$ , where the constant of proportionality depends on  $\deg f$ , or else  $f$  has one of the above special forms. This improves the results in [8, 10], by making the bound explicit, with an exponent that is independent of the degree of  $f$ .

We actually establish a more general result, for the case where  $|A|, |B|, |C|$  are not necessarily equal. The threshold bound  $O(n^{11/6})$  is then replaced by a more involved expression in  $|A|, |B|, |C|$ .

Interest in this problem has recently been revived because of its connection to various problems in combinatorial geometry, including problems on distinct distances in several special configurations, and certain variants of the sum-product problem. Several applications of our results to problems of these kinds will be presented. For example, we show that the number of distinct distances between  $n$  points lying on a constant-degree parametric algebraic curve which does not contain a line, in any dimension, is  $\Omega(n^{4/3})$ , extending the result in [20] and improving the bound in [3].

**Keywords.** Combinatorial geometry, incidences, polynomials.

---

\*Work on this paper by Orit E. Raz and Micha Sharir was supported by Grant 892/13 from the Israel Science Foundation. Work by Micha Sharir was also supported by Grant 2012/229 from the U.S.–Israel Binational Science Foundation, by the Israeli Centers of Research Excellence (I-CORE) program (Center No. 4/11), and by the Hermann Minkowski-MINERVA Center for Geometry at Tel Aviv University. Work by József Solymosi was supported by NSERC, ERC-AdG 321104, and OTKA NK 104183 grants.

<sup>†</sup>School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. [oritraz@post.tau.ac.il](mailto:oritraz@post.tau.ac.il)

<sup>‡</sup>School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. [michas@post.tau.ac.il](mailto:michas@post.tau.ac.il)

<sup>§</sup>Department of Mathematics, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. [solymosi@math.ubc.ca](mailto:solymosi@math.ubc.ca)

# 1 Introduction

## 1.1 Background

In 2000, Elekes and Rónyai [8] considered the following problem. Let  $A, B$  be two sets, each of  $n$  real numbers, and let  $f$  be a real bivariate polynomial of some constant degree. They showed that if  $|f(A \times B)| \leq cn$ , for some constant  $c$  that depends on  $\deg f$ , and for  $n \geq n_0(c)$ , for sufficiently large threshold  $n_0(c)$  that depends on  $c$ , then  $f$  must be of one of the special forms  $f(u, v) = h(\varphi(u) + \psi(v))$ , or  $f(u, v) = h(\varphi(u) \cdot \psi(v))$ , for some univariate polynomials  $\varphi, \psi, h$  over  $\mathbb{R}$ .

In a variant of this setup, we are given, in addition to  $A, B$  and  $f$ , another set  $C$  of  $n$  real numbers, and the quantity  $|f(A \times B)|$  is replaced by the number  $M$  of triples  $(a, b, c) \in A \times B \times C$  such that  $c = f(a, b)$ . Elekes and Rónyai have shown that if  $M = \Omega(n^2)$  then  $f$  must have one of the above special forms. Elekes and Szabó [10] were able to extend this theorem to implicit surfaces  $F(x, y, z) = 0$ , and also showed that, unless  $F$  has a certain specific special form (see [10] for precise formulation and more details), the surface can only contain  $O(n^{2-\eta})$  points of  $A \times B \times C$ , for some exponential ‘gap’  $\eta > 0$  that depends on the degree of the polynomial  $F$  (they do not make the values of  $\eta$  explicit, and point out that it is ‘rather small’). The study of Elekes and Szabó also considers more involved setups, where  $A, B, C$ , and  $F$  are embedded in higher dimensions, and/or the underlying field is the complex field  $\mathbb{C}$ .

In this paper, we prove that, unless  $f$  has one of the aforementioned special forms,  $M = O(n^{11/6})$  (where the constant of proportionality depends on  $\deg f$ ). In doing so, we give two alternative proofs of this result, which we believe to be simpler than the ones in [8, 10]. Our result improves the previous ones, by making the bound on  $M$  explicit, with an exponent that is independent of the degree of  $f$ . (The previous gap is only given in [10]; the former paper [8] only shows that  $M = o(n^2)$ .)

We actually establish a more general result than that of [8], for the case where  $|A|, |B|, |C|$  are not necessarily equal. The threshold bound  $O(n^{11/6})$  is then replaced by a more involved expression in  $|A|, |B|, |C|$  (see Theorem 2 below). This generalization requires a more careful and somewhat more involved analysis. Schwartz, Solymosi and de Zeeuw [23] have recently considered the special ‘unbalanced’ case where  $|A| = |C| = n$  and  $|B| = n^{1/2+\varepsilon}$ , for any fixed  $\varepsilon > 0$ , and showed that the graph of  $f$  must contain  $o(n^{3/2+\varepsilon})$  points of  $A \times B \times C$ , unless  $f$  is of one of the special forms. Our analysis applies in this setup, and slightly improves (and makes more concrete) the bound just mentioned.

The technique used in this paper has some common features with the one used in [8]. A discussion of the similarities and differences between the two approaches is given in the concluding section (Section 6).

Besides being an interesting problem in itself, the Elekes-Rónyai setup, and certain generalizations thereof, such as those considered by Elekes and Szabó [10], arise in many problems in combinatorial geometry. This connection has resurfaced in several recent works, including problems on distinct distances in several special configurations (see Sharir et al. [25] and Pach and de Zeeuw [20] for ad-hoc treatments of these instances). In many of these problems it is essential to allow the sets  $A, B$ , and  $C$  under consideration to be of different sizes (usually, the interest is then in estimating the cardinality of one of these

sets), and then it useful to have the unbalanced version in Theorem 2. To demonstrate this (beyond the refinement of the aforementioned result in [23]), let us describe one of these problems in more detail; other applications will be given in Section 5.

**Distinct distances between two lines.** Consider the following special instance of the distinct distances problem of Erdős. Let  $\ell_1, \ell_2$  be two lines in the plane which are neither parallel nor orthogonal, and let  $P_i$  be a finite set of points on  $\ell_i$ , for  $i = 1, 2$ . Sharir et al. [25] have recently shown that the number of distinct distances between pairs in  $P_1 \times P_2$  is

$$\Omega\left(\min\left\{|P_1|^{2/3}|P_2|^{2/3}, |P_1|^2, |P_2|^2\right\}\right).$$

To see the connection with the Elekes-Rónyai setup, let  $D$  denote the set of all squared distinct distances determined by  $P_1 \times P_2$ , and consider the function  $F : \ell_1 \times \ell_2 \rightarrow \mathbb{R}$ , given by  $F(p, q) = \|p - q\|^2$ . Let  $M$  denote the number of triples  $(p, q, d) \in P_1 \times P_2 \times D$ , for which  $d = F(p, q)$ . By the definition of  $D$ , we have  $M = |P_1||P_2|$ . Thus an upper bound on  $M$  (in terms of  $|P_1|, |P_2|$ , and  $|D|$ ) would yield a lower bound on  $|D|$ . This is essentially the setup in Theorem 2, if we regard  $\ell_1$  and  $\ell_2$  as two copies of  $\mathbb{R}$ , so that  $F$  becomes a quadratic bivariate polynomial over  $\mathbb{R}$ . Then, by Theorem 2, stated below, either  $f$  is of one of the special forms specified in the theorem, which can be shown not to be the case, or else

$$|P_1||P_2| = M = O\left(\left(|P_1|^{2/3}|P_2|^{2/3} + |P_1| + |P_2|\right)|D|^{1/2}\right),$$

which implies that

$$|D| = \Omega\left(\min\left\{|P_1|^{2/3}|P_2|^{2/3}, |P_1|^2, |P_2|^2\right\}\right), \tag{1}$$

which is exactly the bound obtained in [25]. (There are several simple ways to show that  $f$  is not of one of the specific forms, which we omit in this quick discussion.)

**Extensions.** The high-level approach used in this paper can be viewed as an instance of a more general technique, applicable to geometric problems that involve an interaction between three sets of real numbers, where the interaction can be expressed by a general trivariate (constant-degree) polynomial equation  $F(x, y, z) = 0$ . This is very much related to the setup considered by Elekes and Szabó [10], and we will discuss the issues involved in this extension at the end of the paper. Such an extension would facilitate further applications of the new machinery, to a variety of problems of this kind.

Two recent studies, by Solymosi and Sharir [26] and by Raz et al. [21], involve problems of this form. The former paper studies the problem of obtaining a lower bound on the number of distinct distances between three non-collinear points and  $n$  other points in the plane (the lower bound obtained there is  $\Omega(n^{6/11})$ ). The latter paper reconsiders the problem, previously studied by Elekes et al. [9], of obtaining an upper bound on the number of triple intersection points between three families of  $n$  unit circles, where all the circles of the same family pass through a fixed point in the plane (the upper bound obtained there is  $O(n^{11/6})$ ). In both cases the analysis follows a general paradigm, similar to the one in this paper (except that the underlying polynomial is trivariate rather than bivariate), and faces a technical issue that is handled by problem-specific ad-hoc techniques. This issue, which we do not yet spell out, will become clear after digesting our analysis, and will be discussed in the concluding section.

## 1.2 Our results

Our main result, for the case  $|A| = |B| = |C|$ , is as follows.

**Theorem 1** *Let  $A, B$ , and  $C$  be three finite sets of real numbers, each of cardinality  $n$ . Let  $f \in \mathbb{R}[u, v]$  be a bivariate real polynomial of constant degree  $d \geq 2$ , and let  $M$  denote the number of intersection points of the surface  $w = f(u, v)$  with  $A \times B \times C$  in  $\mathbb{R}^3$ . Then either  $M = O(n^{11/6})$ , where the constant of proportionality depends on  $\deg f$ , or  $f$  is of one of the forms  $f(u, v) = h(\varphi(u) + \psi(v))$ , or  $f(u, v) = h(\varphi(u) \cdot \psi(v))$ , for some univariate polynomials  $\varphi, \psi, h$  over  $\mathbb{R}$ .*

As mentioned earlier, in certain applications, the sets  $A, B, C$  are not of the same cardinality. As promised, our analysis caters to these asymmetric situations too, and establishes the following more general result.

**Theorem 2** *Let  $A, B$ , and  $C$  be three finite sets of real numbers. Let  $f \in \mathbb{R}[u, v]$  be a bivariate real polynomial of constant degree  $d \geq 2$ , and let  $M$  denote the number of intersection points of the surface  $w = f(u, v)$  with  $A \times B \times C$  in  $\mathbb{R}^3$ . Then either*

$$M = O\left(\min\left\{|A|^{2/3}|B|^{1/2}|C|^{2/3} + |A|^{1/2}|B|^{2/3}|C|^{2/3} + |A| + |B|, \right. \right. \\ \left. \left. \left(|A|^{2/3}|B|^{2/3} + |A| + |B|\right)|C|^{1/2}\right\}\right),$$

*again where the constant of proportionality depends on  $\deg f$ , or  $f$  is of one of the forms  $f(u, v) = h(\varphi(u) + \psi(v))$ , or  $f(u, v) = h(\varphi(u) \cdot \psi(v))$ , for some univariate polynomials  $\varphi, \psi, h$  over  $\mathbb{R}$ .*

The following is an immediate consequence of the second part of the bound in Theorem 2, which suffices for many of our applications. It is obtained by putting  $C = f(A, B)$  and  $M = |A||B|$ . We do expect, though, that further applications will need to exploit the full generality of our bounds.

**Corollary 3** *Let  $A, B \subset \mathbb{R}$  be two finite sets, and let  $f$  be a bivariate constant-degree real polynomial. Then, unless  $f$  is of one of the special forms specified in the statement of Theorem 2, we have*

$$|f(A, B)| = \Omega\left(\min\left\{|A|^{2/3}|B|^{2/3}, |A|^2, |B|^2\right\}\right).$$

We prove only Theorem 2, and do it in two parts, respectively establishing the first expression (in Section 3) and the second expression (in Section 4) in the asserted bound. (Either of these proofs in itself suffices to obtain Theorem 1 in the balanced case  $|A| = |B| = |C|$ , so there is no need to digest both proofs for this special case, but they provide different bounds and cater to different ranges of  $|A|, |B|$ , and  $|C|$  in the unbalanced case.)

In Section 5 we present several applications of our result. Most of these problems have already been considered in the literature, but our machinery yields improved bounds, and simplifies some of the earlier proofs. These applications include (i) lower bounds on the number of distinct slopes determined by points on a curve, and on the number of distinct distances determined by such points, and (ii) lower bounds for variants of the sum-product problem.

## 2 Preliminaries

**Algebraic preliminaries.** Let  $K$  be a field, and let  $p$  be a bivariate polynomial with coefficients in  $K$ . We say that  $p$  is *decomposable* over  $K$  if we can write  $p(u, v) = r \circ q(u, v) = r(q(u, v))$ , where  $r$  is a univariate polynomial of degree at least two, and  $q$  is a bivariate polynomial, both with coefficients in  $K$ . Otherwise,  $p$  is said to be *indecomposable* (over  $K$ ). It is easy to see that a decomposable polynomial  $p$  over  $K$  is reducible over  $\bar{K}$ , where  $\bar{K}$  stands for the algebraic closure of  $K$ . Indeed, if  $p = r \circ q$ , where  $r$  and  $q$  are as before, then  $p(u, v) = \prod_i (q(u, v) - z_i)$ , where  $z_i, i = 1, \dots, \deg r$ , are the roots of  $r$  (which is indeed a non-trivial factorization since  $\deg r \geq 2$ ).

The following theorem of Stein [28] is crucial for our analysis. (See Shen [27] for another recent application of Stein's theorem to a related problem.) It is concerned with the connection between the decomposability of  $p$  and the reducibility of  $p - \lambda$ , for elements  $\lambda \in \bar{K}$ .

**Theorem 4 (Stein [28])** *Let  $\bar{K}$  be an algebraically closed field, and let  $p$  be a bivariate polynomial with coefficients in  $\bar{K}$ . If  $p$  is indecomposable over  $\bar{K}$ , then*

$$|\{\lambda \in \bar{K} \mid p - \lambda \text{ is irreducible over } \bar{K}\}| < \deg p.$$

The requirement in Theorem 4, that the field under consideration be algebraically closed, is not essential, as shown by the following theorem, taken from Ayad [1, Theorems 4 and 7].

**Theorem 5 (Ayad [1])** *Let  $K$  a field of characteristic zero, and let  $p$  be a bivariate polynomial with coefficients in  $K$ . Then  $p$  is decomposable over  $K$  if and only if it is decomposable over  $\bar{K}$ .*

Combining Theorem 4 and Theorem 5, we obtain the following corollary, which is formulated specifically for our needs in the proof of Theorem 1.

**Corollary 6** *Let  $p$  be a bivariate polynomial in  $\mathbb{R}[x, y]$ . If  $p$  is indecomposable over  $\mathbb{R}$ , then*

$$|\{\lambda \in \mathbb{R} \mid p - \lambda \text{ is irreducible over } \mathbb{R}\}| < \deg p.$$

We also make use of the classical bivariate Bézout's theorem (see, e.g., [4]), again specialized to real polynomials.

**Theorem 7 (Bézout)** *Let  $f$  and  $g$  be two bivariate polynomials over  $\mathbb{R}$ , with degrees  $d_f$  and  $d_g$ , respectively. If  $f$  and  $g$  vanish simultaneously at more than  $d_f d_g$  points of  $\mathbb{R}^2$ , then  $f$  and  $g$  have a common non-trivial factor.*

The following result is useful in analyzing the zero set of a bivariate polynomial on a grid. It is a specialization to two dimensions of the more general result presented in [22] and [34].

**Lemma 8 (Schwartz-Zippel Lemma [22, 34])** *Let  $g$  be a real bivariate polynomial of degree  $\delta$ , and let  $U, V$  be two finite point sets in  $\mathbb{R}^2$ , with  $|U| = |V| = n$ . Then  $g$  has at most  $\delta n$  zeros in  $U \times V$ . In case  $|U| \neq |V|$ , this number is  $\min\{\delta|U|, \delta|V|\}$ .*

**Combinatorial preliminaries.** One of the main ingredients of the proof of Theorem 2 will be a reduction to a problem involving incidences between points and curves in the plane. We therefore recall some basic results in incidence theory, which has its roots in the following classical result of Szemerédi and Trotter [31].

**Theorem 9 (Szemerédi-Trotter [31])** *The number of incidences between  $m$  distinct points and  $n$  distinct lines in  $\mathbb{R}^2$  is  $O(m^{2/3}n^{2/3} + m + n)$ .*

Theorem 9 has seen a number of generalizations. For example, we have:

**Theorem 10 (Pach-Sharir [19])** *Let  $P$  be a collection of  $m$  distinct points in  $\mathbb{R}^2$  and  $S$  a collection of  $n$  distinct curves with  $k$  degrees of freedom, i.e., there exists a constant  $C_0$  such that any two curves can meet in at most  $C_0$  points and at most  $C_0$  curves can contain any  $k$  given points. Then the number of incidences between the points of  $P$  and the curves of  $S$  is  $O\left(m^{\frac{k}{2k-1}}n^{\frac{2k-2}{2k-1}} + m + n\right)$ , where the implicit constant depends only on  $C_0$  and  $k$ .*

(In the Szemerédi-Trotter setup,  $k = 2$ .) Theorem 10 was proved using the Crossing Lemma of Ajtai et al. and of Leighton (see, e.g., [18] for a more recent exposition), which provides a lower bound for the edge-crossing number for graphs embedded in the plane. It was first employed in incidence geometry by Székely [30], where, among other results, it has yielded a simple and elegant proof of the Szemerédi-Trotter theorem.

**Theorem 11 (Crossing Lemma)** *Let  $G = (V, E)$  be a simple graph drawn in the plane. Then*

$$|E| = O\left(|V| + |V|^{2/3}\text{Cr}(G)^{1/3}\right),$$

where  $\text{Cr}(G)$  is the number of pairs  $(e, e')$  of edges of  $E$ , such that the drawing of  $e$  and  $e'$  cross each other.

### 3 Proof of Theorem 2: Part 1

The proof is given in two installments, each establishing (when  $f$  does not have one of the special forms) a different upper bound on  $M$ ; the combination of these bounds yields Theorem 2. The first part is presented in this section, and the second part in Section 4. As noted, when  $|A| = |B| = |C|$ , both parts of the proof yield the same bound, and there is no need to have both.

**Proposition 12** *Let  $A$ ,  $B$ , and  $C$  be three finite sets of real numbers. Let  $f \in \mathbb{R}[u, v]$  be a bivariate real polynomial of constant degree  $d \geq 2$ , and let  $M$  denote the number of intersection points of the surface  $w = f(u, v)$  with  $A \times B \times C$  in  $\mathbb{R}^3$ . Then either*

$$M = O\left(|A|^{2/3}|B|^{1/2}|C|^{2/3} + |A|^{1/2}|B|^{2/3}|C|^{2/3} + |A| + |B|\right),$$

with a constant of proportionality that depends on  $d$ , or  $f$  is of one of the forms  $f(u, v) = h(\varphi(u) + \psi(v))$ , or  $f(u, v) = h(\varphi(u) \cdot \psi(v))$ , for some real univariate polynomials  $\varphi, \psi, h$ .

**Proof.** First consider the case where  $f$  is indecomposable. We put  $d_u = \deg_u(f)$  and  $d_v = \deg_v(f)$  (so  $d \leq d_u + d_v$ ). With each pair  $(a, c) \in A \times C$ , we associate a curve  $\bar{\gamma}_{a,c}$  in  $\mathbb{R}^3$ , defined as

$$\bar{\gamma}_{a,c} := \{(x, y, z) \in \mathbb{R}^3 \mid y = f(a, z) \wedge c = f(x, z)\}. \quad (2)$$

$\bar{\gamma}_{a,c}$  is the intersection curve of the two cylindrical surfaces

$$\sigma_a := \{(x, y, z) \in \mathbb{R}^3 \mid y = f(a, z)\} \quad \text{and} \quad \sigma_c^* := \{(x, y, z) \in \mathbb{R}^3 \mid c = f(x, z)\}$$

in  $\mathbb{R}^3$ . To see that this is indeed a (one-dimensional) curve, note that for every value of  $z$ ,  $y$  is determined uniquely by the equation  $y = f(a, z)$ , and, with the possible exception of at most  $d$  values of  $z$ , there are at most  $d$  values of  $x$  for which  $c = f(x, z)$ . The exceptional values  $z_0$  of  $z$  are those for which  $f(u, z_0)$  is independent of  $u$ , and this can happen at most  $d$  times, because for every such  $z_0$ ,  $z - z_0$  is a factor of  $f(u, z)$ . (If this happens for every value of  $z$  then  $f$  is independent of  $u$  and thus has a degenerate version of the asserted special forms.) Hence the intersection  $\bar{\gamma}_{a,c}$  cannot be two dimensional.

Note that there are at most  $d_u$  pairs  $(a, c) \in A \times C$  that are associated with the same curve. Indeed, let  $\bar{\gamma}$  be some curve of the form (2), and let  $(x, y, z)$  be a generic point of  $\bar{\gamma}$ . A pair  $(a, c)$  which is associated with  $\bar{\gamma}$  satisfies  $y = f(a, z)$  and  $c = f(x, z)$ . This clearly determines  $c$  uniquely, and  $a$  is one of the at most  $d_u$  roots of  $y = f(a, z)$ , regarded as a polynomial in  $a$ . (As just argued, we may assume that at a generic point,  $f(a, z)$  is not a constant polynomial in  $a$ .)

We let  $\gamma_{a,c}$  denote the projection of  $\bar{\gamma}_{a,c}$  onto the  $xy$ -plane in  $\mathbb{R}^3$ , which we identify with  $\mathbb{R}^2$ . In other words,  $\gamma_{a,c}$  is the locus of all points  $(x, y) \in \mathbb{R}^2$  for which there exists  $z \in \mathbb{R}$ , such that  $y = f(a, z)$  and  $c = f(x, z)$ .

Let  $\Gamma := \{\gamma_{a,c} \mid (a, c) \in A \times C\}$  denote the multiset of these curves, allowing for the possibility that the same projection might be shared by more than one original curve, even though the original curves themselves are (up to a constant multiplicity) distinct, as argued above, and let  $I$  denote the number of incidences between the curves of  $\Gamma$  and the points of  $\Pi := A \times C$ ; since the curves of  $\Gamma$  can potentially overlap or coincide, we count incidences with multiplicity: A point lying on  $k$  coinciding curves (or, more precisely, on an irreducible component shared by  $k$  of the curves) contributes  $k$  to the count  $I$ .

Recall that  $M$ , as defined in the theorem, is the number of intersection points of the surface<sup>1</sup>  $y = f(x, z)$  with the point set  $A \times C \times B$  in  $\mathbb{R}^3$ . We obtain an upper bound on  $M$  as follows. For each  $b \in B$ , put

$$\Pi_b = (A \times C)_b := \{(a, c) \in A \times C \mid c = f(a, b)\},$$

and put  $M_b = |\Pi_b| = |(A \times C)_b|$ . We clearly have  $M = \sum_{b \in B} M_b$ .

Fix  $b \in B$ , and note that for any pair of pairs  $(a_1, c_1), (a_2, c_2) \in \Pi_b$ , we have  $(a_1, c_1) \in \gamma_{a_2, c_1}$  and  $(a_2, c_1) \in \gamma_{a_1, c_2}$ . Moreover, for a fixed pair  $(a_1, c_1), (a_2, c_2)$  of this kind, the number of values  $b$  for which  $(a_1, c_1)$  and  $(a_2, c_2)$  both belong to  $\Pi_b$  is at most the constant  $v$ -degree  $d_v$  of  $f$ , unless  $w - f(u, v)$  vanishes identically on the two lines  $(a_1, c_1) \times \mathbb{R}, (a_2, c_2) \times \mathbb{R}$ . The latter situation arises when  $f(a_1, v)$  and  $f(a_2, v)$  are both independent of  $v$ . Similar to what has been argued earlier, this can occur for at most  $d_u^2$  quadruples  $(a_1, c_1), (a_2, c_2)$  of

<sup>1</sup>Note that the roles of the  $y$ -axis and the  $z$ -axis are reversed in the present setup.

this kind (that is, at most  $d_u^2$  pairs  $a_1, a_2$ ; the values of  $c_1, c_2$  are unique in this case, for a fixed choice of  $a_1, a_2$ ). It then follows, using the Cauchy-Schwarz inequality, that

$$\begin{aligned} M &= \sum_{b \in B} M_b \leq \left( \sum_{b \in B} M_b^2 \right)^{1/2} \cdot |B|^{1/2} \\ &\leq (d_v I + d_u^2 |B|)^{1/2} |B|^{1/2} = O\left(I^{1/2} |B|^{1/2} + |B|\right). \end{aligned} \quad (3)$$

Hence deriving an upper bound on  $I$  would yield an upper bound on  $M$ . Bounding  $I$  is an instance of a fairly standard point-curve incidence problem, which can in principle be tackled using the well established machinery reviewed in Section 2. However, to apply this machinery, it is essential for the curves of  $\Gamma$  to have a constant bound on their multiplicity. More precisely, we need to know that no more than  $O(1)$  curves of  $\Gamma$  can share a common irreducible component. When this is indeed the case, we derive an upper bound on the number of incidences, using the following proposition, whose proof is deferred to Section 3.1.

**Proposition 13** *Let  $\Gamma$  and  $\Pi$  be as above, and assume that no more than  $m_0 := d_u d_v + d_u d(d + d_v - 1)$  curves of  $\Gamma$  can share an irreducible component. Then the number  $I$  of incidences between  $\Gamma$  and  $\Pi$  is  $O(|\Gamma|^{2/3} |\Pi|^{2/3} + |\Gamma| + |\Pi|)$ , where the constant of proportionality depends on  $d$ .*

Since  $|\Pi| = |\Gamma| = |A||C|$ , it follows that in this case  $I = O(|A|^{4/3} |C|^{4/3})$ . Plugging this bound into (3), we get  $M = O(|A|^{2/3} |B|^{1/2} |C|^{2/3} + |B|)$ .

In the complementary case, namely when there exist  $m > m_0$  curves of  $\Gamma$  that share an irreducible curve, we start all over again, with the roles of the variables  $u, v$  of  $f$  switched. Although the analysis is fully symmetric, we spell out a few details in the interest of clarity. We now associate with each pair  $(b, c) \in B \times C$  a curve  $\bar{\gamma}_{b,c}$  in  $\mathbb{R}^3$ , defined as

$$\bar{\gamma}_{b,c} := \{(x, y, z) \in \mathbb{R}^3 \mid y = f(z, b) \wedge c = f(z, x)\}.$$

We let  $\gamma_{b,c}$  denote the projection of  $\bar{\gamma}_{b,c}$  onto the  $xy$ -plane in  $\mathbb{R}^3$ , which we identify, as above, with  $\mathbb{R}^2$ . Then  $\gamma_{b,c}$  is the locus of all points  $(x, y) \in \mathbb{R}^2$  for which there exists  $z \in \mathbb{R}$ , such that  $y = f(z, b)$  and  $c = f(z, x)$ . We let  $\tilde{\Gamma} := \{\gamma_{b,c} \mid (b, c) \in B \times C\}$  denote the multiset of the projected curves, and let  $\tilde{I}$  denote the number of incidences (again, counted with multiplicity) between the curves of  $\tilde{\Gamma}$  and the points of  $\tilde{\Pi} := B \times C$ .

With this shuffling of coordinates,  $M$  is now the number of intersection points of the surface  $y = f(z, x)$  with the point set  $B \times C \times A$  in  $\mathbb{R}^3$ . If no more than  $\tilde{m}_0 := d_u d_v + d_v d(d + d_u - 1)$  curves of  $\tilde{\Gamma}$  can share a common irreducible component (note that the roles of  $d_u$  and  $d_v$  are switched, as they should be, in the definition of  $\tilde{m}_0$ ), we apply Proposition 13 to  $\tilde{\Gamma}$  and  $\tilde{\Pi}$  and derive an upper bound on  $\tilde{I}$ . The analysis is fully symmetric to the one given above, and yields

$$M = O\left(\tilde{I}^{1/2} |A|^{1/2} + |A|\right) = O\left(|A|^{1/2} |B|^{2/3} |C|^{2/3} + |A|\right).$$

Thus we have proved the following lemma.



**Lemma 14** *If either (i) no irreducible curve is a component of more than  $m_0$  curves of  $\Gamma$ , or (ii) no irreducible curve is a component of more than  $\tilde{m}_0$  curves of  $\tilde{\Gamma}$ , we have*

$$M = O\left(|A|^{2/3}|B|^{1/2}|C|^{2/3} + |A|^{1/2}|B|^{2/3}|C|^{2/3} + |A| + |B|\right).$$

Note that this is the bound asserted in Proposition 12. It thus remains to consider the case where both  $\Gamma$  and  $\tilde{\Gamma}$  contain curves of large multiplicity, in the precise sense formulated in Lemma 14. We show that in this case  $f$  must have one of the special forms asserted in the theorem. (More precisely, since we are still under the assumption that  $f$  is indecomposable, the analysis yields a more restricted representation of  $f$ ; see below for more detail.) The following proposition “almost” brings us to those forms.

**Proposition 15** *Suppose that there exists an irreducible algebraic curve in  $\mathbb{R}^2$  that is shared by more than  $m_0 = d_u d_v + d_u d(d + d_v - 1)$  distinct curves of  $\Gamma$ . Then  $f$  is of the form*

$$f(u, v) = up(u)q(v) + r(v), \quad (4)$$

for some real univariate polynomials  $p, q, r$ .

The proof of Proposition 15 is given in Section 3.2; it will exploit our (temporary) assumption that  $f$  is indecomposable. Applying a symmetric version of Proposition 15, in which the roles of  $A$  and  $B$ , and the respective  $x$ - and  $z$ -coordinates, are switched, we conclude that we also have

$$f(u, v) = v\tilde{p}(v)\tilde{q}(u) + \tilde{r}(u), \quad (5)$$

for suitable real univariate polynomials  $\tilde{p}, \tilde{q}, \tilde{r}$ .

Equating the two expressions (4) and (5), and substituting  $u = 0$  (resp.,  $v = 0$ ), we get

$$\begin{aligned} r(v) &= v\tilde{p}(v)\tilde{q}(0) + \tilde{r}(0) \\ \tilde{r}(u) &= up(u)q(0) + r(0). \end{aligned}$$

That is,

$$up(u)q(v) + v\tilde{p}(v)\tilde{q}(0) + \tilde{r}(0) = v\tilde{p}(v)\tilde{q}(u) + up(u)q(0) + r(0),$$

or

$$up(u)(q(v) - q(0)) + \tilde{r}(0) = v\tilde{p}(v)(\tilde{q}(u) - \tilde{q}(0)) + r(0),$$

We note that  $r(0) = \tilde{r}(0)$ , because all the other terms in this equation are divisible by  $uv$ . That is, we have

$$up(u)(q(v) - q(0)) = v\tilde{p}(v)(\tilde{q}(u) - \tilde{q}(0)).$$

Assume first that  $q(v) - q(0)$  is not identically zero; that is,  $q$  is not a constant. The equality just derived allows us to write (with a suitable “shift” of the constants of proportionality).

$$up(u) = \tilde{q}(u) - \tilde{q}(0), \quad \text{and} \quad v\tilde{p}(v) = q(v) - q(0).$$

That is, we have,

$$\begin{aligned} f(u, v) &= up(u)q(v) + r(v) = up(u)q(v) + v\tilde{p}(v)\tilde{q}(0) + \tilde{r}(0) \\ &= (\tilde{q}(u) - \tilde{q}(0))q(v) + (q(v) - q(0))\tilde{q}(0) + r(0) = \tilde{q}(u)q(v) + r(0) - \tilde{q}(0)q(0). \end{aligned}$$

In the other case,  $q(v)$  is a constant  $c_0$ , so (4) yields  $f(u, v) = c_0up(u) + r(v)$ .

That is, we have shown that  $f$  is of one of the forms  $\varphi(u) + \psi(v)$  or (up to an additive constant)  $\varphi(u) \cdot \psi(v)$ , for suitable univariate polynomials  $\varphi, \psi$ .

Finally, consider the case where  $f$  is decomposable. Then we may write  $f(u, v) = h(f_0(u, v))$ , where  $f_0$  is an indecomposable bivariate polynomial over  $\mathbb{R}$ , and  $h$  is a (non-linear) univariate polynomial over  $\mathbb{R}$ . We let  $C_0 := h^{-1}(C)$  denote the pre-image of  $C$  under  $h$ . Note that since  $h$  is a polynomial of degree at most  $d$  (actually, at most  $d/2$ ), every  $c \in C$  has at most  $d$  values  $c' \in \mathbb{R}$  for which  $h(c') = c$ . Thus,  $|C_0| \leq d|C|$ , and the number  $M_0$  of intersections of the surface  $z = f_0(u, v)$  with the point set  $A \times B \times C_0$  in  $\mathbb{R}^3$  is at least  $M$ . By the above analysis, applied to the polynomial  $f_0$  and to the sets  $A, B$  and  $C_0$ , we conclude that either  $M = \Theta(M_0) = O(|A|^{2/3}|B|^{1/2}|C|^{2/3} + |A|^{1/2}|B|^{2/3}|C|^{2/3} + |A| + |B|)$ , or  $f_0$  is of one of the two forms  $\varphi(u) + \psi(v)$  or  $\varphi(u) \cdot \psi(v)$  (the extra additive term that we got in the latter case can be “transferred” to the expression defining  $h$ ). Hence, either  $f$  is of one of the two forms  $h(\varphi(u) + \psi(v))$  or  $h(\varphi(u) \cdot \psi(v))$ , or  $M$  satisfies the bound in Proposition 12, as asserted. This finally concludes the proof of the proposition.  $\square$

### 3.1 Proof of Proposition 13

We apply Székely’s technique [30], which is based on the Crossing Lemma, as formulated in Theorem 11 (see also [18, p. 231]). As noted, this is also the approach used in the proof of Theorem 10 in Pach and Sharir [19], but the possible overlap of curves requires some extra (and more explicit) care in the application of the technique.<sup>2</sup>

We begin by constructing a plane embedding of a multigraph  $G$ , whose vertices are the points of  $\Pi$ , and each of whose edges connects a pair  $\pi_1 = (\xi_1, \eta_1)$ ,  $\pi_2 = (\xi_2, \eta_2)$  of points that lie on the same curve  $\gamma_{a,c}$  and are consecutive along (some connected component of)  $\gamma_{a,c}$ ; the edge is drawn along the portion of the curve between the points. One edge for each such curve (connecting  $\pi_1$  and  $\pi_2$ ) is generated, even when the curves coincide or overlap. Thus there might potentially be many edges of  $G$  connecting the same pair of points, whose drawings coincide. Nevertheless, by assumption, the number of overlap at any specific arc is at most  $m_0$ .

In spite of this control on the number of mutually overlapping (or, rather, coinciding) edges, we still face the potential problem that the edge multiplicity in  $G$  (over all curves, overlapping or not, that connect the same pair of vertices) may not be bounded (by a constant). More concretely, we want to avoid edges  $(\pi_1, \pi_2)$  whose multiplicity exceeds  $m_0$ . (By what has just been argued, not all drawings of such an edge can coincide.)

To handle this situation, we observe that, by the symmetry of the definition of the curves,  $\pi_1, \pi_2 \in \gamma_{a,c}$  if and only if  $(a, c) \in \gamma_{\xi_1, \eta_1} \cap \gamma_{\xi_2, \eta_2}$ . Hence, if the multiplicity of the edge connecting  $\pi_1$  and  $\pi_2$  is larger than  $m_0$  then the curves  $\gamma_{\xi_1, \eta_1}$  and  $\gamma_{\xi_2, \eta_2}$  intersect in more than  $m_0$  points, and therefore, since each is the zero set of a polynomial of degree  $d$ , and since  $m_0 \geq d^2$ , Bézout’s theorem (Theorem 7) implies that these curves must overlap in a common irreducible component.

Note that, for a given  $(\xi_1, \eta_1)$ , the curve  $\gamma_{\xi_1, \eta_1}$ , having degree  $d$ , has at most  $d$  irreducible

---

<sup>2</sup>The reason why we cannot apply Theorem 10 directly (with  $k = 2$ ) is that it is possible for a pair of points of  $\Pi$  to have a non-constant arbitrarily large number of curves that pass through both of them; see the analysis below.

components, and, by the assumption on  $\Gamma$ , at most  $m_0$  curves share a common irreducible component. That is, each  $(\xi_1, \eta_1)$  has at most  $(m_0 - 1)d$  “problematic” neighbors that we do not want to connect it to; for any other point, the multiplicity of the edge connecting  $(\xi_1, \eta_1)$  with that point is at most  $m_0$ ; more precisely, at most  $m_0$  curves  $\gamma_{a,c}$  pass through both points.

Consider a point  $(\xi_1, \eta_1)$  and one of its bad neighbors  $(\xi_2, \eta_2)$ ; that is, they are points that lie on too many common curves. Let  $\gamma_{a,c}$  be one of the curves along which  $(\xi_1, \eta_1)$  and  $(\xi_2, \eta_2)$  are neighbors.<sup>3</sup> Then, rather than connecting  $(\xi_1, \eta_1)$  to  $(\xi_2, \eta_2)$  along  $\gamma_{a,c}$ , we continue along the curve from  $(\xi_1, \eta_1)$  past  $(\xi_2, \eta_2)$  until we reach a good point for  $(\xi_1, \eta_1)$ , and then connect  $(\xi_1, \eta_1)$  to that point (along  $\gamma_{a,c}$ ). We skip over at most  $(m_0 - 1)d$  points in the process, but now, having applied this “stretching” to each pair of bad neighbors, each of the modified edges has multiplicity at most  $2m_0$  (the factor 2 comes from the fact that a new edge can be obtained by stretching an original edge from either endpoint).

Note that this edge stretching does not always succeed: It will fail when the connected component  $\gamma'$  of  $\gamma_{a,c}$  along which we connect the points contains fewer than  $(m_0 - 1)d + 2$  points of  $\Pi$ , or when  $\gamma'$  is unbounded and there are fewer than  $(m_0 - 1)d$  points of  $\Pi$  between  $\pi_1, \pi_2$ , and an “end” of  $\gamma$ . Still, the number of new edges in  $G$  is at least  $I(\Pi, \Gamma) - \lambda|\Gamma|$ , for a suitable constant  $\lambda$ , where the term  $\lambda|\Gamma|$  accounts for missing edges on connected components of the curves, for the reasons just discussed. By what have just been argued, the number of edges lost on any single component is at most  $O(m_0d)$ , so  $\lambda = O(m_0d^2) = O(1)$ .

The final ingredient needed for this technique is an upper bound on the number of crossings between the (new) edges of  $G$ . Each such crossing is a crossing between two curves of  $\Gamma$ . Even though the two curves might overlap in a common irreducible component (where they have infinitely many intersection points, none of which is a crossing<sup>4</sup>), the number of proper crossings between them is  $O(1)$ , as follows, for example, from the Milnor–Thom theorem (see [17, 32]), or Bézout’s theorem (Theorem 7). Finally, because of the way the drawn edges have been stretched, the edges, even those drawn along the same original curve  $\gamma_{a,c}$ , may now overlap one another, and then a crossing between two curves may be claimed by more than one pair of edges. Nevertheless, since no edge straddles through more than  $(m_0 - 1)d$  points, the number of pairs that claim a specific crossing is  $O(m_0d) = O(1)$ . Hence, we conclude that the total number of edge crossings in  $G$  is  $O(|\Gamma|^2)$ .

We can now apply Theorem 11, and conclude that

$$I(\Pi, \Gamma) - \lambda|\Gamma| = O\left(|\Pi| + |\Pi|^{2/3}|\Gamma|^{2/3}\right),$$

or

$$I(\Pi, \Gamma) = O\left(|\Pi|^{2/3}|\Gamma|^{2/3} + |\Pi| + |\Gamma|\right),$$

with the constant of proportionality depending on  $d$ , as asserted.  $\square$

### 3.2 Proof of Proposition 15

We may assume that  $\gamma'$  does not contain any portion that is contained in a horizontal line, or, since  $\gamma'$  is assumed to be irreducible, that  $\gamma'$  is not a horizontal line.

<sup>3</sup>We make the pessimistic assumption that they are (consecutive) neighbors along all these curves, which of course does not have to be the case.

<sup>4</sup>In particular, overlapping edges drawn along such a component are not considered to cross one another.

For this, first note that it suffices to consider only curves  $\gamma_{a,c}$  for which  $f(a, v)$  is non-constant (regarded as a polynomial in  $v$ ). Indeed, there are at most  $d_u$  values of  $a$ , for which  $f(a, v)$  is independent of  $v$ . By applying the Schwartz-Zippel lemma (Lemma 8) to the polynomials  $(v, w) \mapsto f(a, v) - w$ , for each of these exceptional values  $a$ , it follows that the number of triples  $(a, b, c) \in A \times B \times C$  for which  $a$  is problematic (in the above sense) and  $f(a, b) = c$  is  $d_u \cdot O((|B| + |C|)d_v)$ , which is subsumed in the asserted bound on  $M$ .

Next, consider a value  $z \in \mathbb{R}$  for which  $f(u, z)$  is independent of  $u$ . There are at most  $d_v$  such values of  $z$ , and each of them determines exactly one value  $c = f(u, z)$  (which depends only on  $z$ ). By applying the Schwartz-Zippel lemma, this time to the polynomials  $(u, v) \mapsto f(u, v) - c$ , for each of these at most  $d_v$  special values  $c$ , it follows that the number of triples  $(a, b, c) \in A \times B \times C$ , for which  $c$  is one of these exceptional values and  $f(a, b) = c$ , is  $d_v \cdot O((|A| + |B|)d)$ . This is again subsumed in the asserted bound on  $M$ , and we may ignore such values  $c \in C$  in our analysis. To recap, by trimming  $A$  and  $C$  accordingly, we may now assume that (i) for each  $a \in A$ ,  $f(a, z)$  is non-constant in  $z$ , and (ii) for each  $c \in C$ , no value  $z_0 \in \mathbb{R}$  yields a constant polynomial  $f(u, z_0)$  (i.e., independent of  $u$ ) whose value is  $c$ .

Now if  $\gamma'$  were a horizontal line of the form  $y = \eta_0$  then, for any curve  $\gamma_{a,c}$  that contains  $\gamma'$ , the system

$$\begin{aligned} f(a, z) &= \eta_0 \\ f(\xi, z) &= c, \end{aligned}$$

in the variables  $\xi, z$ , would have infinitely many solutions. By our assumption,  $f(a, z)$  is non-constant in the variable  $z$ , and hence  $z$  is one of the at most  $d_v$  roots of  $f(a, z) = \eta_0$ . Hence, to get infinitely many solutions  $\xi, z$ , it must be that  $\xi \mapsto f(\xi, z)$  is independent of  $\xi$ . But then  $z$  is one of the exceptional values discussed above, and our pruning of  $C$  ensures that  $f(\xi, z) \notin C$ , and hence  $f(\xi, z) = c$  does not have infinitely many solutions. That is, in the reduced configuration,  $\gamma'$  cannot be a horizontal line.

Let  $(\xi, \eta)$  be a regular point of  $\gamma'$ , and let  $(\alpha, \beta)$  denote the direction vector of the line tangent to  $\gamma'$  at  $(\xi, \eta)$ . For reasons to be clarified later, we choose, as we may, the point  $(\xi, \eta)$  so that  $f_u(\xi, v)$  (regarded as a polynomial in  $\mathbb{R}[v]$ ) is non-constant, and so that the polynomial  $\eta - f(u, v) \in \mathbb{R}[u, v]$  is irreducible over  $\mathbb{R}$ . Indeed, for the former property we only need to avoid the at most  $d_u$  (common) zeros  $\xi$  of the coefficients of the nonconstant monomials of  $f_u$  (regarding  $f_u$  as a polynomial in  $v$ ). For the latter property, we use our assumption that  $f$  is indecomposable. In this case Corollary 6 says that there are at most a constant number of values  $\eta$  for which  $\eta - f(u, v)$  is reducible over  $\mathbb{R}$ . Hence, in total, since  $\gamma'$  does not contain any horizontal line, we need to avoid at most a constant number of points  $(\xi, \eta)$  on  $\gamma'$  to have these two properties, and we let  $(\xi, \eta)$  be one of the other (infinitely many) regular points of  $\gamma'$ .

By assumption, there are  $m > d_u d_v + d_u d(d + d_v - 1)$  pairs  $(a_i, c_i)$ ,  $i = 1, \dots, m$ , such that the curves  $\gamma_{a_i, c_i}$  all contain  $\gamma'$  (and in particular, a neighborhood of  $(\xi, \eta)$  along  $\gamma'$ ). We recall the definitions, for the convenience of the reader.

$$\bar{\gamma}_{a_i, c_i} := \{(x, y, z) \in \mathbb{R}^3 \mid y = f(a_i, z) \wedge c_i = f(x, z)\} = \sigma_{a_i} \cap \sigma_{c_i}^*,$$

where

$$\sigma_{a_i} := \{(x, y, z) \in \mathbb{R}^3 \mid y = f(a_i, z)\}$$

$$\sigma_{c_i}^* := \{(x, y, z) \in \mathbb{R}^3 \mid c_i = f(x, z)\},$$

and

$$\gamma_{a_i, c_i} := \{(x, y) \in \mathbb{R}^2 \mid \exists z \in \mathbb{R} \text{ such that } y = f(a_i, z) \wedge c_i = f(x, z)\}.$$

Then, for each  $i = 1, \dots, m$ , there exists a point  $z_i \in \mathbb{R}$  for which  $p_i := (\xi, \eta, z_i) \in \bar{\gamma}_{a_i, c_i}$ . Observe that the values  $z_i$ ,  $i = 1, \dots, m$ , are not necessarily distinct, but nevertheless the cardinality  $m'$  of  $\{z_i \mid i = 1, \dots, m\}$  is at least  $m/d_u$ . Indeed, for a given value  $z_0$ , the equation  $c = f(\xi, z_0)$  determines  $c$  uniquely, and the equation  $\eta = f(a, z_0)$  determines at most  $d_u$  possible values of  $a$ . (For the latter claim, we note that  $f(a, z_0)$  cannot be independent of  $a$  and satisfy  $f(a, z_0) \equiv \eta$ , for that would imply that  $\eta - f(u, v)$  is divisible by  $v - z_0$ , contradicting the assumption that  $\eta - f(u, v)$  is irreducible and that  $d \geq 2$ .) Then there are at most  $d_u$  pairs  $(a_i, c_i)$  with  $(\xi, \eta, z_0) \in \bar{\gamma}_{a_i, c_i}$ , and hence at most  $d_u$  indices  $i$  for which  $z_i = z_0$ . Thus  $|\{z_i \mid i = 1, \dots, m\}| \geq m/d_u$ , as claimed. Therefore we may assume, by re-indexing if needed, that the values  $z_1, \dots, z_{m'}$  are distinct, with  $m' > d_v + d(d + d_v - 1)$ .

Observe that for at least  $m' - d_v$  of the indices  $1 \leq i \leq m'$ , the point  $p_i$  is a regular point of both  $\sigma_{a_i}, \sigma_{c_i}^*$ . Indeed,  $\sigma_{a_i}$  is a smooth surface since it is the (cylindrical) graph of the polynomial function  $y = f(a_i, z)$ .  $p_i$  is singular in  $\sigma_{c_i}^*$  only if  $f_u(\xi, z_i) = 0$ , but this equation is satisfied by at most  $d_v$  values  $z_i$ . (Recall that by our choice of  $\xi$ , the polynomial  $f_u(\xi, z)$  is non-constant in  $z$ .) Therefore we may assume, by re-indexing if needed, that each of the points  $p_1, \dots, p_{m''}$  is regular on both  $\sigma_{a_i}$  and  $\sigma_{c_i}^*$ , with  $m'' > d(d + d_v - 1)$ .

Let  $\pi_{a_i}, \pi_{c_i}^*$  be the tangent planes of  $\sigma_{a_i}, \sigma_{c_i}^*$  at  $p_i$ , which are now well defined. The normal vectors of  $\pi_{a_i}, \pi_{c_i}^*$  at  $(\xi, \eta, z_i)$  are

$$\mathbf{n}_{a_i} = (0, 1, -f_v(a_i, z_i)), \quad \mathbf{n}_{c_i}^* = (f_u(\xi, z_i), 0, f_v(\xi, z_i)),$$

respectively. Note that these values imply that  $\pi_{a_i} \neq \pi_{c_i}^*$ , and thus the intersection  $\pi_{a_i} \cap \pi_{c_i}^*$  is a line  $l$ . The direction vector of  $l$  is orthogonal to both  $\mathbf{n}_{a_i}, \mathbf{n}_{c_i}^*$ , and is thus given by

$$\mathbf{n}_{a_i} \times \mathbf{n}_{c_i}^* = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \mathbf{n}_{a_i, x} & \mathbf{n}_{a_i, y} & \mathbf{n}_{a_i, z} \\ \mathbf{n}_{c_i, x}^* & \mathbf{n}_{c_i, y}^* & \mathbf{n}_{c_i, z}^* \end{vmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & -f_v(a_i, z_i) \\ f_u(\xi, z_i) & 0 & f_v(\xi, z_i) \end{vmatrix} = \begin{pmatrix} f_v(\xi, z_i) \\ -f_v(a_i, z_i)f_u(\xi, z_i) \\ -f_u(\xi, z_i) \end{pmatrix}.$$

By the assumption on  $(\xi, \eta)$ , the projection of  $\mathbf{n}_{a_i} \times \mathbf{n}_{c_i}^*$  onto the  $xy$ -plane is parallel to  $(\alpha, \beta)$  (recall that  $(\alpha, \beta)$  depends only on  $\gamma'$  and not on a specific choice of  $(a_i, c_i)$ ). That is, for every  $i = 1, \dots, m''$ , we have

$$\beta f_v(\xi, z_i) + \alpha f_v(a_i, z_i) f_u(\xi, z_i) = 0.$$

Consider the system of equations

$$\begin{cases} g_1(a, z) := \beta f_v(\xi, z) + \alpha f_v(a, z) f_u(\xi, z) = 0 \\ g_2(a, z) := \eta - f(a, z) = 0, \end{cases} \quad (6)$$

with  $a, z$  being the unknowns. That is, (6) is satisfied by the  $m'' > d(d + d_v - 1)$  distinct pairs  $(a_i, z_i)$ ,  $i = 1, \dots, m''$ , so it has at least  $d(d + d_v - 1) + 1$  solutions. Since  $\deg g_1 \leq d + d_v - 1$  and  $\deg g_2 = d$ , Bézout's theorem (Theorem 7) implies that the polynomials  $g_1(a, z)$  and  $g_2(a, z)$  in  $\mathbb{R}[a, z]$  must have a (non-constant) common factor. Recalling that, by our choice of  $\eta$ ,  $g_2$  is irreducible over  $\mathbb{R}$ , it follows that  $g_2$  divides  $g_1$ .

Note that the variable  $a$  has the same degree in both  $g_1$  and  $g_2$ . Indeed, its degree in  $g_2$  is  $d_u$  and its degree in  $g_1$  is at most  $d_u$ ; if the latter degree were smaller than  $d_u$ ,  $g_2$  could not divide  $g_1$ . Hence  $g_1$  must be of the form

$$g_1(a, z) \equiv g_2(a, z)h(z), \quad (7)$$

with  $h$  being independent of  $a$ . We write

$$f(u, v) = \sum_{k=0}^{d_u} c_k(v)u^k,$$

so  $f_v(u, v) = \sum_{k=0}^{d_u} c'_k(v)u^k$ . Then (7) becomes

$$\beta f_v(\xi, z) + \alpha \left( \sum_{k=0}^{d_u} c'_k(z)a^k \right) f_u(\xi, z) \equiv \left( \eta - \sum_{k=0}^{d_u} c_k(z)a^k \right) h(z),$$

or

$$\beta f_v(\xi, z) + \alpha f_u(\xi, z)c'_0(z) + \sum_{k=1}^{d_u} \alpha f_u(\xi, z)c'_k(z)a^k \equiv \eta h(z) - h(z)c_0(z) - \sum_{k=1}^{d_u} h(z)c_k(z)a^k.$$

Hence we have, in particular,

$$c'_k(z) \equiv -\frac{h(z)}{\alpha f_u(\xi, z)} c_k(z),$$

for every  $1 \leq k \leq d_u$ . Hence, for every pair of distinct indices  $1 \leq k, l \leq d_u$ , we have

$$c'_k(z)c_l(z) \equiv c_k(z)c'_l(z),$$

or, when  $c_l(z)$  not identically zero,  $\left(\frac{c_k(z)}{c_l(z)}\right)' \equiv 0$ , or  $\frac{c_k(z)}{c_l(z)} \equiv \beta_{kl}$ , for  $\beta_{kl}$  a constant. That is, there exist constants  $\lambda_k \in \mathbb{R}$  and a polynomial  $q(z)$  (independent of  $k$ ), such that  $c_k(z) \equiv \lambda_k q(z)$ , for  $k = 1, \dots, d_u$ . (This also takes care of coefficients  $c_k(z)$  that are identically zero.) Hence,

$$f(u, v) = c_0(v) + uq(v) \sum_{k=0}^{d_u-1} \lambda_{k+1} u^k.$$

Putting  $p(u) := \sum_{k=0}^{d_u-1} \lambda_{k+1} u^k$ , we conclude that  $f$  is of the form  $f(u, v) = c_0(v) + up(u)q(v)$ , as asserted.  $\square$

## 4 Proof of Theorem 2: Part 2

So far we have considered two reductions where the parametric plane in which point-curve incidences have been analyzed contained  $A \times C$  and  $B \times C$ , respectively. In this section we consider a somewhat more natural (or “standard”) setup, in which the dependence on  $z$  is eliminated right away, and the emphasis is mainly on the sets  $A$  and  $B$ . This approach leads to the second upper bound in Theorem 2. That is, we show:

**Proposition 16** *Let  $A$ ,  $B$ , and  $C$  be three finite sets of real numbers. Let  $f \in \mathbb{R}[u, v]$  be a bivariate real polynomial of constant degree  $d \geq 2$ , and let  $M$  denote the number of intersection points of the surface  $w = f(u, v)$  with  $A \times B \times C$  in  $\mathbb{R}^3$ . Then either*

$$M = O\left(\left(|A|^{2/3}|B|^{2/3} + |A| + |B|\right)|C|^{1/2}\right),$$

*with a constant of proportionality that depends on  $d$ , or  $f$  is of one of the forms  $f(u, v) = h(\varphi(u) + \psi(v))$ , or  $f(u, v) = h(\varphi(u) \cdot \psi(v))$ , for some real univariate polynomials  $\varphi, \psi, h$ .*

**Proof.** Arguing as in Section 3, we may assume, without loss of generality, that  $f$  is indecomposable. In the present “standard” setup, the curves are defined by

$$\gamma_{a,b} = \{(x, y) \mid f(a, x) = f(b, y)\}, \quad (8)$$

for  $a, b \in A$ . We let  $\Gamma$  denote the multiset of these curves (so  $|\Gamma| = |A|^2$ , counting curves with multiplicity), and put  $\Pi := B^2$ . For each  $c \in C$  let  $M_c$  denote the number of pairs  $(a, b) \in A \times B$  such that  $f(a, b) = c$ . Then  $M = \sum_{c \in C} M_c$ . Let  $I = I(\Pi, \Gamma)$  denote the number of incidences between the points of  $\Pi$  and the curves of  $\Gamma$ . A similar argument to that in the preceding section shows that, for every pair of pairs  $(a_1, b_1), (a_2, b_2) \in A \times B$  with  $f(a_1, b_1) = f(a_2, b_2)$ , we have  $(b_1, b_2) \in \gamma_{a_1, a_2}$ . This implies, as before, that  $I \geq \sum_{c \in C} M_c^2$ , and then

$$M = \sum_{c \in C} M_c \leq \left(\sum_{c \in C} M_c^2\right)^{1/2} \cdot |C|^{1/2} \leq I^{1/2} |C|^{1/2}. \quad (9)$$

Hence the problem is reduced to obtaining an upper bound on  $I$ .

**Bounding the number of incidences.** Again, we are concerned with situations where many curves of  $\Gamma$  share a common irreducible component  $\gamma'$ . We want to show that in this case  $f$  must have one of the special forms asserted in the theorem. The complementary case, in which no component  $\gamma'$  is shared by too many curves, will lead to the incidence bound that we are after, as we detail next.

Note that in the definition of the curves (8), the roles of the variables  $u, v$  of  $f$  are symmetric and can be reversed. Namely, we can consider the “dual” curves<sup>5</sup>

$$\gamma_{\xi, \eta}^* = \{(x, y) \mid f(x, \xi) = f(y, \eta)\}, \quad (10)$$

for  $\xi, \eta \in B$ . Then, for  $(a_1, b_1), (a_2, b_2) \in A \times B$ , we have that  $(b_1, b_2) \in \gamma_{a_1, a_2}$  if and only if  $(a_1, a_2) \in \gamma_{b_1, b_2}^*$ . We let  $\Gamma^*$  denote the multiset of the curves  $\gamma_{\xi, \eta}^*$ , with  $(\xi, \eta) \in B^2$ , and put  $\Pi^* := A^2$ .

Let  $\Gamma_0$  (resp.,  $\Gamma_0^*$ ) denote the set of irreducible curves that are shared by more than  $m_0 := \max\{d_u^2, d_v^2\}$  curves of  $\Gamma$  (resp., of  $\Gamma^*$ ). Incidences between points  $(\xi, \eta) \in \Pi$  and curves  $\gamma_{a,b} \in \Gamma$ , such that the portion of  $\gamma_{a,b}$  containing  $(\xi, \eta)$  is not in  $\Gamma_0$ , and the portion of  $\gamma_{\xi, \eta}^*$  containing  $(a, b)$  is not in  $\Gamma_0^*$ , can be analyzed via Székely’s crossing-lemma technique (see Theorem 11), as we did in the proof of Proposition 13 in Section 3, and their number is thus at most

$$O\left(|\Pi|^{2/3}|\Gamma|^{2/3} + |\Pi| + |\Gamma|\right) = O\left(|A|^{4/3}|B|^{4/3} + |A|^2 + |B|^2\right), \quad (11)$$

---

<sup>5</sup>It is interesting to observe that, in contrast, in the setup of Section 3 the dual scenario involves the same kind of curves as the primal one.

where the constant of proportionality depends on  $d$ . (In more details, the only difference from the proof of Proposition 13 is in analyzing the multiplicity of the edges in the constructed graph  $G$ ; this multiplicity can be interpreted as the number of dual curves that share an irreducible factor, and hence is bounded by  $m_0$ , due to our exclusion of incidences that occur on curves of  $\Gamma_0^*$ .)

If at least half the quadruples in  $Q := \{(a, \xi, b, \eta) \in (A \times B)^2 \mid f(a, \xi) = f(b, \eta)\}$  correspond to incidences that occur on primal curves that are not in  $\Gamma_0$ , and on dual curves that are not in  $\Gamma_0^*$ , then the expression in (11) serves as the desired upper bound on  $I$ , which, combined with (9), yields the bound on  $M$  asserted in the proposition. We may thus assume that at least half the quadruples in  $Q$  correspond to incidences that occur either on primal curves of  $\Gamma_0$  or on dual curves of  $\Gamma_0^*$ , and, without loss of generality (by the symmetry of the two settings), that at least a quarter of the quadruples in  $Q$  correspond to incidences that occur on primal curves of  $\Gamma_0$ . We refer to the curves in  $\Gamma_0$  as *popular curves*.

**Curves with larger multiplicity: Reducibility and its consequences.** Consider then incidences that occur on popular curves. Let  $Q_0$  denote the subset of quadruples  $(a, p, b, q)$  in  $Q$  such that the irreducible component of  $\gamma_{a,b}$  that contains  $(p, q)$  is popular. We have the following key proposition, whose proof is given in Section 4.1.

**Proposition 17** *There exists a bivariate polynomial  $h$ , that depends only on  $f$ , such that  $\deg h \leq 2d^2$ , and  $h$  is of one of the forms  $\varphi(u) + \psi(v)$ , or  $\varphi(u) \cdot \psi(v)$ , for some univariate polynomials  $\varphi, \psi$  (that depend only on  $f$ ), and such that the following property holds. For at least  $|Q_0| - O(|A||B|)$  quadruples  $(a, p, b, q)$  of  $Q_0$ , we have  $h(a, p) = h(b, q)$ .*

Recall that we are currently assuming that at least a quarter of the quadruples in  $Q$  correspond to incidences that occur on curves of  $\Gamma_0$ . That is,  $|Q_0| = \Theta(|Q|)$ . We remove from  $C$  all elements  $c$  for which  $c - f(u, v)$  is reducible. Since we continue to assume that  $f$  is indecomposable, Corollary 6 tells us that the number of values  $c$  for which  $c - f(u, v)$  is reducible is smaller than  $d$ , and an application of the Schwartz-Zippel lemma (Lemma 8) implies that, for each such  $c$ , the number of pairs  $(a, p) \in A \times B$  satisfying  $f(a, p) = c$  is  $O((|A| + |B|)d)$ , so the number of quadruples associated with these exceptional values is  $O((|A| + |B|)^2)$ , and we simply ignore them in our analysis (recall that this argument has already been used in earlier parts of the analysis). We may assume that  $|Q_0|$  is much larger than this bound, for otherwise  $|Q|$  satisfies the bound in (11) and we are done. By removing at most  $O(|A||B|)$  additional quadruples from  $Q_0$ , as prescribed in Proposition 17, we may also assume that  $h(a, p) = h(b, q)$ , for each surviving quadruple  $(a, p, b, q) \in Q_0$ .

With these reductions, we conclude that there exists a pair  $(a, p) \in A \times B$  that participates in at least  $t := |Q_0|/(|A||B|)$  quadruples  $(a, p, b, q)$  of  $Q_0$ , so that, for  $c = f(a, p) = f(b, q)$ ,  $c - f(u, v)$  is irreducible, and  $c' = h(a, p) = h(b, q)$ . We put  $c' = h(a, p)$ , and note that  $c$  and  $c'$  are fixed once  $(a, p)$  is fixed.

Assume first that the polynomial  $h(u, v)$  is indecomposable. In this case we can apply a fully symmetric argument to  $h$ , and, by possibly discarding another set of  $O((|A| + |B|)^2)$  quadruples from  $Q_0$ , involving pairs  $(a, p)$  for which  $h(a, p) - h(u, v)$  is reducible, be left with quadruples involving only pairs  $(a, p)$  for which  $(a, p)$  satisfies all the properties assumed so far, and  $h(a, p) - h(u, v)$  is also irreducible.



Now fix a pair  $(a, p) \in A \times B$  satisfying all the above properties. For at least  $t$  pairs  $(b, q)$ , we have

$$\begin{aligned} f(a, p) - f(b, q) &= 0 \\ h(a, p) - h(b, q) &= 0. \end{aligned}$$

That is, the polynomials  $f(a, p) - f(u, v)$  and  $h(a, p) - h(u, v)$  have at least  $t$  common roots. Hence, unless  $t$  is at most some suitable constant (in which case  $|Q| = O(|Q_0|) = O(|A||B|)$ , well below the bound in Proposition 16), these polynomials have a common factor. But since  $f(a, p) - f(u, v)$  and  $h(a, p) - h(u, v)$  are both irreducible, they must be proportional to one another, implying that

$$f(u, v) = \alpha h(u, v) + \beta,$$

for suitable constants  $\alpha, \beta$ .

Next we consider the case where  $h$  is decomposable. We have the following simple claim.

**Claim 18** *Let  $h$  be any polynomial of the form  $h(u, v) = \varphi(u) + \psi(v)$ , for some non-constant univariate real polynomials  $\varphi, \psi$ . Then  $h$  is indecomposable.*

**Proof.** Suppose to the contrary that  $h$  is decomposable, and write  $h(u, v) = r(h_0(u, v))$ , for some nonlinear univariate polynomial  $r$  and a bivariate polynomial  $h_0$ . We have

$$\varphi(u) + \psi(v) = r(h_0(u, v)).$$

Taking derivatives of both sides with respect to the variable  $u$  yields

$$\varphi'(u) = r'(h_0(u, v))(h_0)_u(u, v).$$

Since by assumption  $\varphi'(u) \neq 0$  (and by the unique factorization property over  $\mathbb{R}$ ),  $r'(h_0(u, v))$  must divide  $\varphi'(u)$  and thus be independent of  $v$ . Since  $r$  is nonlinear, this is easily seen to imply that  $h_0$  itself, and thus  $h$  too, must be independent of  $v$ , contradicting the assumption that  $\psi(v)$  is non-constant.  $\square$

Hence, a decomposable  $h$  must be of the form  $h(u, v) = \varphi(u)\psi(v)$ . We write  $h(u, v) = r(h_0(u, v))$ , where  $h_0$  is indecomposable, and  $r$  is a nonlinear univariate real polynomial. Using the same argument as before, we may assume that we are left with quadruples involving only pairs  $(a, p)$  for which all the previous properties are satisfied (with the values  $c, c'$  fixed), and  $h_0(a, p) - h_0(u, v)$  is irreducible. Now consider the equation  $r(s) = c'$ , which has at most  $\deg h \leq d^2$  real roots, and enumerate those roots  $s$  for which  $s - h_0(u, v)$  is irreducible as  $s_1, \dots, s_{d'}$ , for some  $d' \leq d^2$ . We have at least  $t$  elements  $(a, p) \in (A \times B)_c$  for which  $h_0(a, p) \in \{s_1, \dots, s_{d'}\}$ , and so there is an index  $j$  such that at least  $t/d' \geq t/d^2$  elements  $(a, p) \in (A \times B)_c$  satisfy  $h_0(a, p) = s_j$ .

Now fix  $(a, p) \in (A \times B)_c$  as one of these pairs. Then, for at least  $t/d^2$  pairs  $(b, q)$ , which share the same properties with  $(a, p)$ , we have

$$\begin{aligned} f(a, p) - f(b, q) &= 0 \\ h_0(a, p) - h_0(b, q) &= 0. \end{aligned}$$

That is, the two irreducible polynomials  $f(a, p) - f(u, v)$  and  $h_0(a, p) - h_0(u, v)$  have at least  $t/d^2$  common roots. Arguing as in the previous case, assuming  $t$  to be sufficiently large, this implies that

$$f(u, v) = \alpha h_0(u, v) + \beta,$$

for suitable constants  $\alpha, \beta$ .

To complete the analysis we claim that  $h_0$  is of the form  $h_0(u, v) = \varphi_1(u)\psi_1(v) + z$ , for some real univariate polynomials  $\varphi_1, \psi_1$ , and  $z \in \mathbb{R}$ . Indeed, we can factor  $r$  over  $\mathbb{R}$  into a product of linear and irreducible quadratic factors. If there is at least one linear factor then the corresponding factor of  $r(h_0(u, v))$ , of the form  $h_0(u, v) - z$ , for some  $z \in \mathbb{R}$ , divides  $\varphi(u)\psi(v)$ , and thus must be of the form  $\varphi_1(u)\psi_1(v)$ , implying the claim. Otherwise, consider an irreducible factor of the form  $(h_0 - w)^2 + z^2$ , for  $z, w \in \mathbb{R}$ . That is, we must have

$$(h_0(u, v) - w)^2 + z^2 = \varphi_1(u)\psi_1(v),$$

for suitable polynomials  $\varphi_1, \psi_1$ . Taking derivatives of both sides of this identity with respect to the variable  $u$ , we get

$$2(h_0(u, v) - w)(h_0)_u(u, v) = \varphi_1'(u)\psi_1(v).$$

Hence  $h_0(u, v) - w$  must divide  $\varphi_1'(u)\psi_1(v)$ , so, as above,  $h_0$  has the asserted form.

In summary, we have covered all subcases, and have shown that the existence of a large number of curves that overlap in a common irreducible component implies that  $f$  has one of the special forms. So far we have assumed that  $f$  is indecomposable, but, as noted in the beginning of the proof, the case where  $f$  is decomposable can be handled as in Section 3. We have thus completed the proof of Proposition 16, that is, of the second part of the proof of Theorem 2.  $\square$

## 4.1 Proof of Proposition 17

We begin the analysis with the following lemma, which derives a useful property involving reducibility of bivariate polynomials of the special form  $p(x) - q(y)$  that we consider here. To prepare for the lemma, we introduce the following notation. For a bivariate polynomial  $u(x, y)$ , let  $u^*(x, y)$  denote the polynomial which is the sum of all monomials of  $u(x, y)$  of maximum total degree. We refer to  $u^*(x, y)$  as the *leading-terms polynomial*, or *LT-polynomial* in short, of  $u(x, y)$ . Note that if  $u(x, y) = v(x, y)w(x, y)$  then necessarily we also have  $u^*(x, y) = v^*(x, y)w^*(x, y)$ .

**Lemma 19** *Let  $f, g \in \mathbb{R}[x, y]$  be two polynomials of the special form*

$$\begin{aligned} f(x, y) &= p_1(x) - q_1(y) \\ g(x, y) &= p_2(x) - q_2(y), \end{aligned}$$

*and assume that they have a nontrivial common factor. Assume also that, for  $i = 1, 2$ ,  $\deg p_i = \deg q_i$ , and denote this common value by  $d_i$ . Write*

$$\begin{aligned} [p_1(x) - q_1(y)]^* &= a_1x^{d_1} - b_1y^{d_1} \\ [p_2(x) - q_2(y)]^* &= a_2x^{d_2} - b_2y^{d_2}, \end{aligned}$$

for suitable nonzero coefficients  $a_1, b_1, a_2, b_2$ . Then we have

$$\left(\frac{a_1}{b_1}\right)^{d_2} = \left(\frac{a_2}{b_2}\right)^{d_1}. \quad (12)$$

**Proof.** The proof proceeds by induction on  $\max\{d_1, d_2\}$ . The case where  $d_1 = d_2 = 1$  is easy, because two linear polynomials have a common factor if and only if they are proportional to one another, so we have  $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ , which establishes the induction base.

Consider then the pair  $f(x, y), g(x, y)$  assumed in the lemma, with respective degrees  $d_1, d_2$ , and assume that the lemma is true for all pairs of polynomials of maximum degree smaller than  $\max\{d_1, d_2\}$ .

Let  $r(x, y)$  denote the common factor of  $f$  and  $g$ . Assume without loss of generality that  $d_2 \geq d_1$ . By dividing  $p_1$  and  $q_1$  by  $b_1$  and  $p_2$  and  $q_2$  by  $b_2$ , we may write

$$\begin{aligned} f^*(x, y) &= [p_1(x) - q_1(y)]^* = c_1 x^{d_1} - y^{d_1} \\ g^*(x, y) &= [p_2(x) - q_2(y)]^* = c_2 x^{d_2} - y^{d_2}, \end{aligned}$$

where  $c_1 = a_1/b_1 \neq 0$  and  $c_2 = a_2/b_2 \neq 0$ .

As noted above, the fact that  $r$  is a common factor of  $f$  and  $g$  implies that  $r^*$  is a common factor of  $f^*$  and  $g^*$ . Hence it is also a factor of

$$\begin{aligned} c_2 x^{d_2-d_1} f^* - c_1 g^* &= c_1 y^{d_2} - c_2 x^{d_2-d_1} y^{d_1} \\ &= y^{d_1} \left( c_1 y^{d_2-d_1} - c_2 x^{d_2-d_1} \right). \end{aligned}$$

However,  $r^*$  cannot have any power of  $y$  as a factor because then it could not divide  $f^*$  and  $g^*$ . Hence  $r^*$  is also a factor of

$$h^*(x, y) = c_2 x^{d_2-d_1} - c_1 y^{d_2-d_1}.$$

If  $d_2 = d_1$  then  $h^* = c_2 - c_1$ , which must be 0 since  $r^*$  divides  $h^*$ . That is, in this case we have

$$\frac{a_1}{b_1} = \frac{a_2}{b_2}, \quad \text{or} \quad \left(\frac{a_1}{b_1}\right)^{d_2} = \left(\frac{a_2}{b_2}\right)^{d_1},$$

as asserted. Assume then that  $d_1 < d_2$ . Since  $\max\{d_1, d_2 - d_1\} < \max\{d_1, d_2\}$ , we may apply the induction hypothesis to<sup>6</sup>  $f^*$  and  $h^*$ , and therefore conclude that

$$\left(\frac{c_2}{c_1}\right)^{d_1} = c_1^{d_2-d_1}, \quad \text{or} \quad c_2^{d_1} = c_1^{d_2}, \quad \text{or} \quad \left(\frac{a_1}{b_1}\right)^{d_2} = \left(\frac{a_2}{b_2}\right)^{d_1}.$$

This establishes the induction step and thereby completes the proof.  $\square$

Let us return to the setup under consideration, where we have a multiset  $\Gamma$  of curves of the form

$$\gamma_{a,b} = \{(x, y) \mid f(a, x) = f(b, y)\},$$

for  $a, b \in A$ , and we want to analyze the situation where we have an irreducible component  $\gamma'$  that is contained in at least  $m_0 + 1 = \max\{d_u^2, d_v^2\} + 1$  of these curves. Denote by  $S(\gamma')$  the set of all the pairs  $(a, b) \in A^2$  that define these curves.

<sup>6</sup>Note that here we work directly with the LT-polynomials.

Fix three generic, regular points  $(\xi_1, \eta_1), (\xi_2, \eta_2), (\xi_3, \eta_3) \in \gamma'$ , so that they satisfy conditions (i) and (ii), spelled out shortly below. We have

$$f(a, \xi_i) - f(b, \eta_i) = 0, \quad \text{for } i = 1, 2, 3, \text{ and for } (a, b) \in S(\gamma'). \quad (13)$$

Write  $f(u, v) = \sum_{i=0}^{d_u} c_i(v)u^i$ , and let  $0 \leq k \leq d_u$  denote the maximal index for which  $c_k(v)$  is non-constant in  $v$ . If  $k$  does not exist then  $f(u, v)$  does not depend on  $u$ , so it has (a degenerate version of) one of the special forms in the theorem. If  $k = 0$  then  $f(u, v)$  is of the form  $\varphi(u) + c_0(v)$ , for a suitable univariate polynomial  $\varphi$ , so again it has one of the special forms. Assume then that  $k \geq 1$  and that the points  $(\xi_i, \eta_i)$ ,  $i = 1, 2, 3$ , are chosen so that (i)  $c_k$  does not vanish at any of the six points  $\xi_i, \eta_i$ , and (ii) for any  $1 \leq i < j \leq 3$ ,  $c_k(\xi_i) \neq c_k(\xi_j)$  and  $c_k(\eta_i) \neq c_k(\eta_j)$ . To see that such a choice is possible, we note that  $c_k$  is non-constant and that  $\gamma'$  is not a line parallel to one of the axes. (The latter property holds because the polynomial defining  $\gamma'$  is a factor of  $f(a, x) - f(b, y)$ , and we may assume (arguing as in Section 3) that  $(a, b)$  is not one of the at most  $d_u^2$  pairs for which  $f(a, x) - f(b, y)$  is independent of one of the variables  $x, y$ ). Hence, there are infinitely many ways to choose such points  $(\xi_i, \eta_i)$ . From (13), we get the system of equations

$$\begin{aligned} f(a, \xi_1) - f(a, \xi_2) - f(b, \eta_1) + f(b, \eta_2) &= 0 \\ f(a, \xi_3) - f(b, \eta_3) &= 0, \end{aligned} \quad (14)$$

which, as equations in  $a$  and  $b$ , have at least  $m_0 + 1 > d_u^2$  common roots; the purpose of the first equation in (14) is to get rid of the  $v$ -independent leading  $u$ -terms of  $f$ . Since the first equation in (14) has degree  $k$  and the second has degree  $d_u$ , and since  $m_0 + 1 > d_u^2 \geq d_u k$ , Bézout's theorem (Theorem 7) tells us that

$$f(a, \xi_1) - f(a, \xi_2) - f(b, \eta_1) + f(b, \eta_2), \quad f(a, \xi_3) - f(b, \eta_3)$$

(regarded as polynomials in  $a, b$ ), have a common factor. We can therefore apply Lemma 19 to these two polynomials. We have

$$\begin{aligned} [f(a, \xi_1) - f(a, \xi_2) - f(b, \eta_1) + f(b, \eta_2)]^* &= (c_k(\xi_1) - c_k(\xi_2))a^k - (c_k(\eta_1) - c_k(\eta_2))b^k \\ [f(a, \xi_3) - f(b, \eta_3)]^* &= c_{d_u}(\xi_3)a^{d_u} - c_{d_u}(\eta_3)b^{d_u}, \end{aligned}$$

where, by construction, all the coefficients in the right-hand sides are non-zero. Lemma 19 thus implies that

$$\left( \frac{c_k(\xi_1) - c_k(\xi_2)}{c_k(\eta_1) - c_k(\eta_2)} \right)^{d_u} = \left( \frac{c_{d_u}(\xi_3)}{c_{d_u}(\eta_3)} \right)^k. \quad (15)$$

We distinguish between two cases, depending on whether (15) holds for  $k < d_u$  or  $k = d_u$ .

**The case  $k < d_u$ .** Recall that we assume that  $k \geq 1$ , so we have  $1 \leq k < d_u$ . In this case the right hand-side of (15) is 1. Hence, replacing  $(\xi_1, \eta_1)$  by a generic point<sup>7</sup>  $(\xi, \eta)$  along  $\gamma'$ , we conclude that, for all points  $(\xi, \eta) \in \gamma'$ ,

$$c_k(\eta) - c_k(\xi) = \lambda, \quad (16)$$

<sup>7</sup>Note that we may assume that none of the points of  $\gamma' \cap B^2$  under consideration is an isolated point of  $\gamma'$ . Indeed, each curve of  $\Gamma$  may have at most  $d^2$  such points, and hence the number of incidences between the points of  $\Pi$  and the isolated points of curves of  $\Gamma$ , counted with multiplicity, is at most  $d^2|A|^2$ .

for some fixed parameter  $\lambda$ , which depends on  $\gamma'$ .

In other words, all points on  $\gamma'$  satisfy the system

$$\begin{aligned} f(a, x) - f(b, y) &= 0 \\ c_k(x) - c_k(y) - \lambda &= 0, \end{aligned}$$

for any of the pairs  $(a, b) \in S(\gamma')$ . By Bézout's theorem,  $f(a, x) - f(b, y)$  and  $c_k(x) - c_k(y) - \lambda$  have a common factor.

By Claim 18,  $q(x, y) := c_k(x) - c_k(y)$  is indecomposable. Stein's theorem (see Corollary 6) then implies that  $c_k(x) - c_k(y) - \lambda$  is irreducible for all but fewer than  $d_v$  values of  $\lambda$ . Hence there are fewer than  $d_v$  curves  $\gamma'$ , such that  $\gamma'$  is a component of  $c_k(x) - c_k(y) - \lambda = 0$ , for some constant parameter  $\lambda$ , and  $c_k(x) - c_k(y) - \lambda$  is reducible. Thus we can ignore such curves in our analysis, since their (total) contribution to the number of incidences, and hence to the cardinality of  $Q$ , is  $O(|A||B|)$ , which is subsumed in the bound of (11). Let  $Q_0$  denote the subset of  $Q$  obtained by discarding the  $O(|A||B|)$  quadruples that correspond to these incidences. We may therefore assume that, for the value of  $\lambda$  associated with  $\gamma'$ ,  $c_k(x) - c_k(y) - \lambda$  is indeed irreducible. It follows that  $c_k(x) - c_k(y) - \lambda$  divides  $f(a, x) - f(b, y)$ , for every pair  $(a, b) \in S(\gamma')$ . That is, we may write

$$f(a, x) - f(b, y) \equiv (c_k(x) - c_k(y) - \lambda)g(x, y), \quad (17)$$

for a suitable polynomial  $g$  that depends on  $a$  and  $b$ .

Consider now the symmetric representation  $f(u, v) = \sum_{j=0}^{d_v} \tilde{c}_j(u)v^j$ , and let  $1 \leq \ell \leq d_v$  denote the maximal index for which  $\tilde{c}_\ell(u)$  is non-constant in  $u$ . (As before, we may assume that  $\ell$  exists and that  $\ell \geq 1$ .)

Fix a pair  $(a, b) \in S(\gamma')$ , and substitute  $y = x$  in (17). By discarding at most  $O(|A||B|)$  quadruples of  $Q_0$ , we may assume that  $\tilde{c}_\ell(a) \neq \tilde{c}_\ell(b)$ . Indeed, by the Schwartz-Zippel lemma (Lemma 8), the equation  $\tilde{c}_\ell(x) - \tilde{c}_\ell(y) = 0$  has at most  $O(|A|)$  solutions  $(a, b) \in A^2$ . Then, by another application of the Schwartz-Zippel lemma, for each such  $(a, b)$ ,  $\gamma_{a,b}$  is incident to at most  $O(|B|)$  points of  $B^2$ . Hence the number of quadruples that correspond to incidences occurring on those curves  $\gamma_{a,b}$  is  $O(|A||B|)$ , and we trim  $Q_0$  further, by removing from it all these quadruples. This yields the identity

$$f(a, x) - f(b, x) \equiv -\lambda g(x, x). \quad (18)$$

Note that, by the definition of  $\ell$  and by our assumption that  $\tilde{c}_\ell(a) \neq \tilde{c}_\ell(b)$ , the polynomial on the left-hand side of (18) is of degree exactly  $\ell$ . Let  $g_1(x) := g(x, x)$ . Comparing the degrees of the polynomial on both sides of (18), we get  $\ell = \deg g_1 \leq \deg g = d_v - e < d_v$ , where  $e$  is the degree of  $c_k$ . In particular, the leading term of  $f(a, x) - f(b, y)$ , which is of degree  $d_v$ , is independent of  $a$  and  $b$ .

By (17), we have

$$[f(a, x) - f(b, y)]^* \equiv [c_k(x) - c_k(y)]^* [g(x, y)]^*,$$

or

$$[g(x, y)]^* \equiv \frac{[f(a, x) - f(b, y)]^*}{[c_k(x) - c_k(y)]^*}. \quad (19)$$

As just argued, the numerator (and certainly also the denominator) of the right-hand side of (19) is independent of  $a$  and  $b$ . Hence, up to a (non-zero) multiplicative constant, we have

$$[g(x, y)]^* \equiv \frac{x^{d_v} - y^{d_v}}{x^e - y^e} \equiv x^{d_v - e} + x^{d_v - 2e} y^e + \dots + x^e y^{d_v - 2e} + y^{d_v - e}. \quad (20)$$

In particular, substituting  $y = x$  again,  $[g(x, x)]^*$  becomes  $\frac{d_v}{e} x^{d_v - e} \neq 0$ . Comparing the leading terms on both sides of (18) yields

$$[f(a, x) - f(b, x)]^* \equiv -\lambda [g(x, x)]^*,$$

or

$$\tilde{c}_\ell(a) - \tilde{c}_\ell(b) = -\lambda d_v / e.$$

So, for every  $(x, y) \in \gamma'$ , and every  $(a, b) \in S(\gamma')$ , we have

$$\tilde{c}_\ell(a) - \tilde{c}_\ell(b) = -(d_v / e)(c_k(x) - c_k(y)),$$

or

$$\tilde{c}_\ell(a) + (d_v / e)c_k(x) = \tilde{c}_\ell(b) + (d_v / e)c_k(y). \quad (21)$$

Put

$$h(u, v) := \tilde{c}_\ell(u) + (d_v / e)c_k(v),$$

and observe that  $h$  does not depend on  $\gamma'$ , that  $\deg h \leq d \leq 2d^2$ , and that  $h$  satisfies the property assumed in the proposition.

**The case  $k = d_u$ .** First note that, arguing as in the previous case, and by the symmetry of the two setups,  $\ell < d_v$  implies  $k < d_u$ . Hence, we can assume that in this case we also have  $\ell = d_v$ . We replace  $(\xi_3, \eta_3)$  in (15) by a generic point  $(\xi, \eta)$  along  $\gamma'$ , and conclude that all points  $(\xi, \eta)$  on  $\gamma'$ , except possibly for some finite discrete subset (recall the previous comment concerning isolated points), satisfy

$$c_{d_u}(\eta) = \mu c_{d_u}(\xi), \quad (22)$$

for some fixed parameter  $\mu$ , which depends on  $\gamma'$ . (In contrast,  $c_{d_u}$  itself depends only on  $f$  and not on  $\gamma'$ .)

Consider now the system

$$\begin{aligned} f(a, x) - f(b, y) &= 0 \\ \mu c_{d_u}(x) - c_{d_u}(y) &= 0, \end{aligned}$$

of polynomials in  $x, y$ , which has infinitely many solutions for each pair  $(a, b) \in S(\gamma')$  (they both vanish on  $\gamma'$ ). That is, the polynomials  $f(a, x) - f(b, y)$  and  $\mu c_{d_u}(x) - c_{d_u}(y)$  have a common factor. The corresponding system of leading terms is (after dividing the second equation by the leading coefficient of  $c_{d_u}$ )

$$\begin{aligned} \tilde{c}_{d_v}(a)x^{d_v} - \tilde{c}_{d_v}(b)y^{d_v} \\ \mu x^{e'} - y^{e'}, \end{aligned}$$

for a suitable exponent  $e' \geq 1$ . By discarding at most  $O(|A||B|)$  further quadruples from  $Q_0$ , we may assume that neither of the coefficients  $\tilde{c}_{d_v}(a), \tilde{c}_{d_v}(b)$  is zero. Indeed, the equation

$\tilde{c}_{d_v}(x) = 0$  has at most  $d_u$  solutions in  $\mathbb{R}$ , and hence there are at most  $2d_u|A|$  pairs  $(a, b)$  in  $A^2$  such that one of  $a, b$  is one of these solutions. For each pair  $(a, b)$  of this form, by the Schwartz-Zippel lemma (Lemma 8),  $\gamma_{a,b}$  is incident to at most  $O(|B|)$  points of  $B^2$ . Hence the total number of quadruples involving those curves is at most  $O(|A||B|)$ , and we remove all of them from  $Q_0$ . Then Lemma 19 implies that

$$\left(\frac{\tilde{c}_{d_v}(a)}{\tilde{c}_{d_v}(b)}\right)^{e'} = \mu^{d_v} = \left(\frac{c_{d_u}(y)}{c_{d_u}(x)}\right)^{d_v}.$$

That is,

$$\tilde{c}_{d_v}(a)^{e'} c_{d_u}(x)^{d_v} = \tilde{c}_{d_v}(b)^{e'} c_{d_u}(y)^{d_v}, \quad (23)$$

for each  $(x, y) \in \gamma'$  and each  $(a, b) \in S(\gamma')$ . Put

$$h(u, v) := \tilde{c}_{d_v}(u)^{e'} c_{d_u}(v)^{d_v}$$

and observe that  $h$  does not depend on  $\gamma'$ , that  $\deg h \leq 2d^2$ , and that  $h$  satisfies the property assumed in the proposition.

In either of the two cases, substituting  $(x, y) = (p, q)$  in (21) or in (23) yields  $h(a, p) = h(b, q)$ , as asserted.  $\square$

## 5 Applications

### 5.1 Directions determined by a planar point set

For a finite point set  $P \subset \mathbb{R}^2$  we denote by  $S(P)$  the number of distinct directions determined by pairs of points of  $P$ . The study of sets that determine few distinct directions was initiated by Scott [24]. He conjectured that  $S(P) \geq |P| - 1$  for any non-collinear planar point set. This was settled in the affirmative by Ungar [33]. Sets for which equality holds are called *critical* by Jamison [16] and those with one additional direction, i.e., sets satisfying  $S(P) = |P|$ , are called *near-critical*. Jamison gives an overview of the known critical and near-critical configurations in the Euclidean plane and, among other results, characterizes critical or near-critical configurations that lie in the union of two or three straight lines.

Little is known about plane sets with  $S(P) = |P| + 1$ , let alone  $S(P) \leq c|P|$ , for some constant  $c > 0$ . One result in this direction, due to Elekes [6], is that the Jamison configurations are essentially the only structures that can satisfy even the much weaker requirement  $S(P) \leq c|P|$ , provided that  $P$  contains  $\alpha|P|$  collinear points, where  $\alpha$  is a sufficiently large fraction (see [6] for the exact statement). In the same paper Elekes also proves the following theorem.

**Theorem 20 (Elekes [6])** *Let  $\gamma \subset \mathbb{R}^2$  be a curve of the form  $y = f(x)$ , where  $f$  is some constant-degree polynomial, and  $\deg f \geq 3$ . Then, for any finite point set  $P \subset \gamma$ , we have  $S(P) = \omega(|P|)$ .*

Theorem 2 (more precisely, Corollary 3) yields the following significant sharpening of this result.

**Theorem 21** *Let  $\gamma \subset \mathbb{R}^2$  be a curve of the form  $y = f(x)$ , where  $f$  is some constant-degree polynomial, and  $\deg f \geq 3$ . Then, for any finite point set  $P \subset \gamma$ , we have  $S(P) = \Omega(|P|^{4/3})$ .*

**Proof.** Consider the polynomial function  $g(x, y) := \frac{f(x)-f(y)}{x-y}$ . It is shown in [6] that  $g$  is not of one of the special forms in Theorem 2. The asserted bound then follows from Corollary 3.  $\square$

For completeness, we mention the following recent result of Elekes and Szabó [11].

**Theorem 22 (Elekes and Szabó [11])** *Let  $\gamma \subset \mathbb{R}^2$  be an irreducible constant-degree algebraic curve. Then, either  $\gamma$  is a conic section, or, for any finite point set  $P \subset \gamma$ , we have  $S(P) = \omega(|P|)$ .*

It is easy to construct examples of a finite set  $P$  that lies on a conic section, such as a circle or a pair of lines, that determines only  $\Theta(|P|)$  distinct directions.

For the proof, Elekes and Szabó use their earlier result from [10], also mentioned in the introduction, which deals with implicit surfaces of the form  $F(x, y, z) = 0$ , where  $F$  is some constant-degree trivariate polynomial. Extending Theorem 1 to such surfaces, which we believe is possible (see concluding section for more details), will sharpen the ‘gap’ given by Theorem 22.

## 5.2 Distinct distances: Special configurations

For a finite point set  $P$  (lying in some Euclidean space), we denote by  $D(P)$  the number of distinct distances determined by pairs of points of  $P$ . In [3], Charalambides proved the following theorem.

**Theorem 23 (Charalambides [3])** *Let  $\gamma \subset \mathbb{R}^d$  be a constant-degree irreducible algebraic curve, which is not an algebraic helix. Then, for any finite point set  $P \subset \gamma$ ,  $D(P) = \Omega(|P|^{5/4})$ .*

(Here an *algebraic helix* is either a line, or a curve that, up to a rigid motion, admits a parametrization of the form  $(u_1, \dots, u_k) \mapsto (\alpha_1 \cos u_1, \alpha_1 \sin u_1, \dots, \alpha_k \cos u_k, \alpha_k \sin u_k) \in \mathbb{R}^{2k} \subset \mathbb{R}^d$ , for some parameter  $k \leq d/2$ .)

For the special case  $d = 2$ , Pach and de Zeeuw [20] managed to improve the lower bound obtained in Theorem 23 as follows; their result generalizes the bound (1) mentioned in the introduction.

**Theorem 24 (Pach and de Zeeuw [20])** *Let  $\gamma \subset \mathbb{R}^2$  be a constant-degree irreducible algebraic curve which is not a line or a circle. Then, for any finite point set  $P \subset \gamma$ ,  $D(P) = \Omega(|P|^{4/3})$ .*

Using our machinery, we obtain the same lower bound of  $\Omega(|P|^{4/3})$  for points on a curve in an arbitrary (constant) dimension  $d$ , improving the bound given by Theorem 23; our result however is somewhat weaker, because it requires the curve  $\gamma$  to have a polynomial parameterization.



**Theorem 25** *Let  $\gamma \subset \mathbb{R}^d$  be a curve of the form  $\gamma(t) = (x_1(t), \dots, x_d(t))$ , for  $t \in \mathbb{R}$ , where  $x_1(t), \dots, x_d(t)$  are some constant-degree polynomials. Then, either  $\gamma$  is a line, or, for any finite point set  $P \subset \gamma$ ,  $D(P) = \Omega(|P|^{4/3})$ .*

**Proof.** Consider the bivariate polynomial function

$$f(t, s) := \|\gamma(t) - \gamma(s)\|^2 = \sum_{i=1}^d (x_i(t) - x_i(s))^2, \quad \text{for } t, s \in \mathbb{R}.$$

By shifting the coordinate frame, we may assume that  $x_i(0) = 0$ , and we also assume that  $x_i(t)$  is not identically zero, for each  $i = 1, \dots, d$ ; coordinates for which  $x_i \equiv 0$  do not affect the function  $f$  and can simply be ignored. Suppose that  $f$  is of one of the forms  $h(\varphi(t) + \psi(s))$ , or  $h(\varphi(t) \cdot \psi(s))$ , for some univariate polynomials  $h, \varphi, \psi$ . We claim that in this case  $\gamma$  must be a line (the converse statement is easy to verify).

Consider first the multiplicative special form. That is, assume that

$$\sum_{i=1}^d (x_i(t) - x_i(s))^2 = h(\varphi(t)\psi(s)),$$

for suitable univariate polynomials  $h, \varphi, \psi$ . Substituting  $t = s$ , the left-hand side in the above identity is zero, and hence we must have that  $\varphi(t)\psi(t) \equiv x_0$ , where  $x_0$  is a real root of  $h$ . However, this can occur only if the polynomials  $\varphi, \psi$  are both constants, that is, only if  $h(\varphi(t)\psi(s))$  is a constant independent of  $t$  and  $s$ . Since this quantity corresponds to the distance between two points, represented by the parameters  $t, s$ , on the (one-dimensional) curve  $\gamma$ , this yields a contradiction.

Next consider the additive special form. That is, assume that

$$\sum_{i=1}^d (x_i(t) - x_i(s))^2 = h(\varphi(t) + \psi(s)),$$

for  $h, \varphi, \psi$ , as above. Substituting  $t = s$ , as above, we must have that  $\varphi(t) + \psi(t) \equiv x_0$ , where  $x_0$  is a real root of  $h$ . The last identity then becomes

$$\sum_{i=1}^d (x_i(t) - x_i(s))^2 = h(\varphi(t) - \varphi(s) + x_0). \quad (24)$$

Moreover, the multiplicity of  $x_0$  (as a root of  $h$ ) is at least two (because this is the multiplicity of the factor  $t - s$  of the polynomial on the left-hand side of this identity).

Taking derivatives on both sides of (24) twice, once with respect to  $t$  and then with respect to  $s$ , yields

$$2 \sum_{i=1}^d x'_i(t)x'_i(s) = h''(\varphi(t) - \varphi(s) + x_0)\varphi'(t)\varphi'(s). \quad (25)$$

Comparing the leading terms, we see that  $h''$  must be a constant. Indeed, the leading term of the left-hand side, divided by  $\varphi'(t)\varphi'(s)$ , is of the form  $\alpha t^e s^e$ , for some integer  $e \geq 0$ , and some constant  $\alpha$ . Hence the leading term of  $h''(\varphi(t) - \varphi(s) + x_0)$  must also have this

form. Let  $e'$  denote the degree of  $h$  (as a univariate polynomial). The leading term of  $h''(\varphi(t) - \varphi(s) + x_0)$  (as a bivariate polynomial of  $t$  and  $s$ ) is, up to a constant multiplicative factor, the same as the leading term of  $(\varphi(t) - \varphi(s) + x_0)^{e'}$ . Then clearly, in order to have the form  $\alpha t^e s^e$ , it must be that  $e' = 0$ . Hence  $h''$  is a constant, and thus  $h$  is a polynomial of degree two. Since  $x_0$  is a multiple root of  $h$ , this implies

$$\sum_{i=1}^d (x_i(t) - x_i(s))^2 = (\varphi(t) - \varphi(s))^2. \quad (26)$$

We have assumed that neither of the polynomials  $x_i$  has a constant term, and we may assume this also holds for  $\varphi$ . We then get

$$\sum_{i=1}^d x_i^2(t) + \sum_{i=1}^d x_i^2(s) - 2 \sum_{i=1}^d x_i(t)x_i(s) \equiv \varphi^2(t) + \varphi^2(s) - 2\varphi(t)\varphi(s).$$

This in turn implies that  $\sum_{i=1}^d x_i^2(t) \equiv \varphi^2(t)$  (all the other terms are divisible by  $s$ ), and thus  $\sum_{i=1}^d x_i(t)x_i(s) \equiv \varphi(t)\varphi(s)$ . That is, the scalar product of the two vectors  $(x_1(t), \dots, x_d(t))$  and  $(x_1(s), \dots, x_d(s))$  is equal to the product of their lengths, so these vectors must be parallel, for every pair  $s, t \in \mathbb{R}$ . In other words,  $\gamma$  must be a line through the origin. Removing our assumption that  $x_i(0) = 0$ ,  $\gamma$  can be any line, as claimed.

Hence, if  $\gamma$  is not a line,  $f$  cannot have one of the special forms, and Corollary 3 implies that  $D(P) = \Omega(|P|^{4/3})$ , as asserted.  $\square$

Recently, Bronner et al. [2] considered a bipartite version of the distinct distances problem, where we are given two finite point sets  $P_1, P_2$  in  $\mathbb{R}^d$ , with  $d \geq 3$ , and the points of  $P_1$  are contained in a line  $\ell$  (and without any restriction on the points of  $P_2$ ). They showed that the number of distinct distances spanned by pairs of points from  $P_1 \times P_2$  is

$$\Omega(\min\{|P_1|^{2/3}|P_2|^{2/3}, |P_1|, |P_2|\}),$$

unless many of the points of  $P_2$  lie either on a cylinder, with  $\ell$  as its axis, or on an hyperplane orthogonal to  $\ell$  (see [2] for the exact statement, and for more results of this kind).

### 5.3 Sum-product-type estimates

Variants of the sum-product problem have been studied intensively since the work of Erdős and Szemerédi [13], where it was shown that there exists  $c > 0$  such that for any finite set  $A \subset \mathbb{Z}$ , one has

$$|A + A| + |A \cdot A| \geq |A|^{1+c},$$

where  $A + A = \{u + v \mid u, v \in A\}$ , and  $A \cdot A = \{uv \mid u, v \in A\}$ . Much of the subsequent extensive work aimed either to give an explicit (lower) bound for  $c$ , or to derive generalizations of the sum-product lower bound. The currently best known lower bound is due to Solymosi [29], and asserts that, for any finite set  $A \subset \mathbb{R}$ , one has

$$|A + A| + |A \cdot A| \geq \frac{|A|^{4/3}}{2 \lceil \log |A| \rceil^{1/3}}.$$

One of the significant generalizations of this problem is the work by Elekes et al. [7] who showed that, for any given finite set  $A \subset \mathbb{R}$ , and a strictly convex (or concave) function  $f$  defined on an interval containing  $A$ , one has

$$|A + A| + |f(A) + f(A)| = \Omega(|A|^{5/4}). \quad (27)$$

Elekes et al. actually prove the following more general result.

**Theorem 26 (Elekes et al. [7])** *Let  $A \subset \mathbb{R}$  be a finite set, with  $|A| = N$ , and let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a strictly convex (or concave) function, defined on an interval containing  $A$ . Then, for any pair of sets  $C, D \subset \mathbb{R}$  satisfying  $|C||D| \geq N$ , we have*

$$|A + C||f(A) + D| = \Omega(N^{3/2}(|C||D|)^{1/2}). \quad (28)$$

Substituting  $C = A$ ,  $D = f(A)$  in (28) gives (27) (here the strict convexity or concavity of  $f$  implies that  $|D| = \Omega(|A|)$ ).

Theorem 26 immediately implies the following lemma (part of which was used in Green and Tao [14]).

**Lemma 27** *Let  $A, B \subset \mathbb{R}$  be finite sets, such that  $|B|^2 \geq |A|$ , and let  $f_1, f_2 : \mathbb{R} \rightarrow \mathbb{R}$ . Suppose that there exist  $x_1 < x_2 < \dots < x_c$ , for some constant index  $c \geq 1$ , such that  $f_1$  is strictly concave or convex, and  $f_2$  is strictly monotone (increasing or decreasing), on each open interval  $(x_i, x_{i+1})$ . Then*

$$|A + B| + |f_1(A) + f_2(B)| = \Omega(|A|^{3/4}|B|^{1/2}). \quad (29)$$

**Proof.** Let  $A'$  denote the largest set among the sets  $A_i = A \cap (x_i, x_{i+1})$ , which is of cardinality at least  $|A|/c$ , and similarly let  $B'$  denote the largest set among the sets  $B_i = B \cap (x_i, x_{i+1})$ , which is of cardinality at least  $|B|/c$ . Apply Theorem 26 to  $A'$  (instead of  $A$ ), with  $C = B'$ ,  $f = f_1$ , and  $D = f_2(B')$ . Note that  $|D| = |B'|$ , by the monotonicity of  $f_2$ . This yields

$$|A + B| \cdot |f_1(A) + f_2(B)| \geq |A' + B'| \cdot |f_1(A') + f_2(B')| = \Omega(|A|^{3/2}|B|).$$

Hence one of  $|A + B|$  or  $|f_1(A) + f_2(B)|$  is of cardinality  $\Omega(|A|^{3/4}|B|^{1/2})$ , as asserted.  $\square$

Recently, Shen [27] proved the following generalization of (27).

**Theorem 28 (Shen [27])** *Let  $f$  be a bivariate constant-degree real polynomial. Then, either  $f$  is of the form  $f(x, y) = h(ax + by)$ , for some univariate polynomial  $h$  and constants  $a, b \in \mathbb{R}$ , or, for any finite set  $A \subset \mathbb{R}$ , one has*

$$|A + A| + |f(A, A)| = \Omega(|A|^{5/4}).$$

In view of Corollary 3, Shen's result is interesting only in the case where  $f$  is of one of the special forms from Theorem 2, since in the complementary case we always have  $|f(A, A)| = \Omega(|A|^{4/3})$ . Moreover, for functions  $f$  having one of the special forms, Theorem 28 follows from Lemma 27, as will be shown later. Since the overall conclusion is somewhat asymmetric, let us state it explicitly.

**Corollary 29** *Let  $f$  be a bivariate constant-degree real polynomial. If  $f$  is not of one of the forms  $f(x, y) = h(\varphi(x) + \psi(y))$ , or  $f(x, y) = h(\varphi(x) \cdot \psi(y))$ , for some univariate polynomials  $h, \varphi, \psi$ , then, for any finite set  $A \subset \mathbb{R}$ , one has*

$$|A + A| + |f(A, A)| = \Omega(|A|^{4/3}).$$

*Otherwise, either  $f$  is of the form  $f(x, y) = h(ax + by)$ , for some constants  $a, b \in \mathbb{R}$ , or, for any finite set  $A \subset \mathbb{R}$ , one has*

$$|A + A| + |f(A, A)| = \Omega(|A|^{5/4}).$$

Actually, our analysis implies the following more general result, involving two different sets (of potentially different cardinalities), which does not appear in [27].

**Corollary 30** *Let  $f$  be a bivariate constant-degree real polynomial. If  $f$  is not of one of the forms  $f(x, y) = h(\varphi(x) + \psi(y))$ , or  $f(x, y) = h(\varphi(x) \cdot \psi(y))$ , for some univariate polynomials  $h, \varphi, \psi$ , then, for any pair of finite sets  $A, B \subset \mathbb{R}$ , one has*

$$|A + B| + |f(A, B)| = \Omega\left(|A|^{2/3}|B|^{2/3} + |A| + |B|\right).$$

*Otherwise, either  $f$  is of the form  $f(x, y) = h(ax + by)$ , for some univariate polynomial  $h$  and constants  $a, b \in \mathbb{R}$ , or, for any two finite sets  $A, B \subset \mathbb{R}$ , with  $|A|^{1/2} \leq |B| \leq |A|^2$ , one has*

$$|A + B| + |f(A, B)| = \Omega\left(\min\left\{|A|^{3/4}|B|^{1/2}, |A|^{1/2}|B|^{3/4}\right\}\right).$$

**Proof.** If  $f$  is not of one of the forms  $f(x, y) = h(\varphi(x) + \psi(y))$ , or  $f(x, y) = h(\varphi(x) \cdot \psi(y))$ , Corollary 3 implies

$$|A + B| + |f(A, B)| \geq |f(A, B)| = \Omega\left(|A|^{2/3}|B|^{2/3} + |A| + |B|\right).$$

To treat the complementary case, assume that  $f$  has one of the above forms, and that  $\varphi$ , say, is a nonlinear polynomial (the special form of  $f$  asserted in Theorem 28 is merely an equivalent way of saying that  $h$  has the additive form and that both  $\varphi$  and  $\psi$  are linear). We may assume that  $\psi$  is not a constant, for otherwise  $f$  is independent of  $y$  and thus can be written as  $h(ax + by)$ . If  $f$  has the form  $f(x, y) = h(\varphi(x) + \psi(y))$ , we apply Lemma 27, letting  $f_1 = \varphi$  and  $f_2 = \psi$ ; since  $f_1$  is a nonlinear constant-degree polynomial, and  $f_2$  is a constant-degree polynomial (which is non-constant), it is easy to see that the conditions of the lemma are satisfied. If  $f$  is of the multiplicative special form  $f(x, y) = h(\varphi(x)\psi(y))$ , we again apply Lemma 27, this time letting  $f_1 = \ln \circ \varphi$  and  $f_2 = \ln \circ \psi$ . In either of the two cases, the lemma implies the asserted bound.  $\square$

## 6 Conclusion

At a high level, there are some common features of the analysis of Elekes and Rónyai [8] and ours, but there are considerable differences in the actual analysis (and results). At the risk of making the comparison somewhat informal and imprecise, we list a few similarities and differences.

(i) Both techniques double count “quadruples”, or rather “quintuples” of various kinds. For example, in our second proof of Theorem 2 we consider quintuples  $(a, b, p, q, c) \in A^2 \times B^2 \times C$  such that  $f(a, p) = f(b, q) = c$  (the parameter  $c$  is implicit in our setup). In contrast, Elekes and Rónyai consider quintuples of the form  $(a, b_1, b_2, c_1, c_2) \in A \times B^2 \times C^2$ , such that  $f(a, b_1) = c_1$  and  $f(a, b_2) = c_2$ .

(ii) In both cases the quintuples are interpreted as incidences between points and curves in a suitable parametric plane. The reductions are different, though, and the remainders of the proofs are a consequence of the parameterization. Elekes and Rónyai obtain curves of the form  $\{(f(t, b_i), f(t, b_j)) \mid b_i, b_j \in B\}$ , which are rationally or polynomially parameterizable. Our curves are different (and the curves appearing in the two proofs of the theorem are also different from one another).

(iii) One notable difference is that Elekes and Rónyai’s goal was only to establish a dichotomy between the case where (in our notation)  $M$  is quadratic and  $f$  has one of the special forms, and the complementary case. They did not set up to obtain a concrete ‘gap’ between the two cases, as we do in this paper. (Such a (weaker) gap has been obtained later, by Elekes and Szabó [10], in their treatment of a more general setup.)

(iv) Both proofs use rather elementary algebra of polynomials, of different sorts.

An obvious direction for further research is to extend the machinery developed in this paper to the more general setup of Elekes and Szabó [10], involving the vanishing of a trivariate polynomial  $F(x, y, z)$  on a three-dimensional grid  $A \times B \times C$ . The general high-level approach is clear: We can consider the set  $Q$  of quadruples  $(a, a', b, b') \in A^2 \times B^2$ , such that there exists  $c \in C$  satisfying  $F(a, b, c) = F(a', b', c) = 0$  (compared with what has just been noted, this actually counts the quintuples  $(a, b, a', b', c)$ ), relate the number  $|Q|$  of such quadruples, via the Cauchy-Schwarz inequality, to  $M$ , the number of zeros of  $F$  on the grid, and then interpret each quadruple as an incidence between, e.g., the point  $(b, b')$  and a suitable curve  $\gamma_{a, a'}$ , defined in analogy with the curves of Section 4. Again, the main technical hurdle is to handle situations where too many of these curves (or of their duals, flipping the roles of  $A$  and  $B$ ) overlap in a common irreducible component. That is, the challenge is to show that if this is not the case then  $Q$  can be bounded via a standard incidence bound, as we did above, and then the bound  $M = O(n^{11/6})$  would follow, and if there exist overlaps of large multiplicity, then  $F$  must be special, e.g., in the sense of [10].

We believe that our analysis can also be applied over the complex field, and leave this extension as (what we hope would be an easy) open problem.

Another interesting challenge is to extend the result to higher-dimensional grids; see Schwartz et al. [23] for an initial attempt in this direction for four-dimensional grids. An even more challenging direction would be to extend the analysis to cases where the constituent sets  $A, B, C$  of the grid are not one-dimensional. In these cases the problem would translate to incidences between points and higher-dimensional varieties, typically, points and two-dimensional varieties in  $\mathbb{R}^4$  (where  $A$  and  $B$  are sets of points in the plane).

Another interesting project is to obtain a sharp calibration of the dependence of the bounds in this paper on the degree of  $f(x, y)$ . For example, our results and those of [20], show that the number of distinct distances between  $n$  points on a constant-degree curve (which is neither a line or a circle) in the plane is  $\Omega(n^{4/3})$ . On the other hand, for any set of  $n$  points in the plane there exists a curve of degree  $d = O(\sqrt{n})$  that passes through all

the points (e.g., see [15]), and then the nearly linear upper bound on the number of distinct distances in the grid construction of Erdős [12] suggests that we will not be able to prove a superlinear lower bound when  $d = \Theta(\sqrt{n})$ . Is there any hope in deriving a lower bound that depends on  $d$ , and interpolate between the two extreme situations noted above?

Another open problem is to improve the bound on  $M$  in Theorems 1 and 2. We are not aware of any non-trivial lower bound for  $M$ , and suspect it to be much smaller.

Finally, it would be interesting to find additional applications of the results of this paper.

## References

- [1] M. Ayad, Sur les polynômes  $f(X, Y)$  tels que  $K[f]$  est intégralement fermé dans  $K[X, Y]$ , *Acta Arithmetica* 105 (2002), 9–28.
- [2] A. Bronner, M. Sharir and A. Sheffer, Distinct distances on a line and a curve, manuscript, 2014.
- [3] M. Charalambides, Exponent gaps on curves via rigidity, in arXiv:1307.0870 (2013).
- [4] D. A. Cox, J. Little and D. O’Shea, *Using Algebraic Geometry*, Springer-Verlag, 2nd Edition, Heidelberg 2005.
- [5] G. Elekes, Sums versus products in number theory, algebra and Erdős geometry—A survey, in *Paul Erdős and his Mathematics II*, Bolyai Math. Soc., Stud. 11, Budapest, 2002, pp. 241–290.
- [6] G. Elekes, On linear Combinatorics III, *Combinatorica*, 19(1) (1999), 43–53.
- [7] G. Elekes, M. Nathanson and I. Ruzsa, Convexity and sumsets, *J. Number Theory* 83(2) (2000), 194–201.
- [8] G. Elekes and L. Rónyai, A combinatorial problem on polynomials and rational functions, *J. Combinat. Theory Ser. A* 89 (2000), 1–20.
- [9] G. Elekes, M. Simonovits and E. Szabó, A combinatorial distinction between unit circles and straight lines: How many coincidences can they have? *Combinat. Probab. Comput.* 18 (2009), 691–705.
- [10] G. Elekes and E. Szabó, How to find groups? (and how to use them in Erdős geometry?), *Combinatorica* 32 (2012), 537–571.
- [11] G. Elekes and E. Szabó, On triple lines and cubic curves: The Orchard Problem revisited, in arXiv:1302.5777 (2013).
- [12] P. Erdős, On a set of distances of  $n$  points, *Amer. Math. Monthly* 53 (1946), 248–250.
- [13] P. Erdős and E. Szemerédi, On sums and products of integers, in: P. Erdős, L. Alpár, G. Halász, and A. Sárközy, editors, *Studies in Pure Mathematics, To the Memory of Paul Turán*, 1983, Birkhäuser Verlag, Basel, pp. 213–218.
- [14] B. Green and T. Tao, On sets defining few ordinary lines, in arXiv:1208.4714 (2012).

- [15] L. Guth and N. H. Katz, Algebraic methods in discrete analogs of the Kakeya problem, *Advances Math.* 225 (2010), 2828–2839. Also in arXiv:0812.1043v1.
- [16] R. E. Jamison, Planar configurations which determine few slopes, *Geometriae Dedicata* 16 (1984), 17–34.
- [17] J. Milnor, On the Betti numbers of real varieties, *Proc. Amer. Math. Soc.*, 15(2) (1964), 275–280.
- [18] J. Pach and P. K. Agarwal, *Combinatorial Geometry*, Wiley-Interscience, New York 1995.
- [19] J. Pach and M. Sharir, On the number of incidences between points and curves, *Combinat. Probab. Comput.* 7 (1998), 121–127.
- [20] J. Pach and F. de Zeeuw, Distinct distances on algebraic curves in the plane, in arXiv:1308.0177 (2013).
- [21] O. E. Raz, M. Sharir, and J. Solymosi, On triple intersections of three families of unit circles, manuscript, 2013.
- [22] J. Schwartz, Fast probabilistic algorithms for verification of polynomial identities, *J. ACM* 27(4) (1980), 701–717.
- [23] R. Schwartz, J. Solymosi, and F. de Zeeuw, Extensions of a result of Elekes and Rónyai, *J. Combinat. Theory, Ser. A* 120(7) (2013), 1695–1713.
- [24] P. R. Scott. On the sets of directions determined by  $n$  points, *Amer. Math. Monthly* 77 (1970), 502–505.
- [25] M. Sharir, A. Sheffer, and J. Solymosi, Distinct distances on two lines, *J. Combinat. Theory, Ser. A* 20 (2013), 1732–1736.
- [26] M. Sharir and J. Solymosi, Distinct distances from three points, in arXiv:1308.0814 (2013).
- [27] C. Shen, Algebraic methods in sum-product phenomena, *Israel J. Math.* 188(1) (2012), 123–130.
- [28] Y. Stein, The total reducibility order of a polynomial in two variables, *Israel J. Math.* 68(1) (1989), 109–122.
- [29] J. Solymosi, Bounding multiplicative energy by the sumset, *Advances Math.* 222 (2009), 402–408.
- [30] L. Székely, Crossing numbers and hard Erdős problems in discrete geometry, *Combinat. Probab. Comput.* 6 (1997), 353–358.
- [31] E. Szemerédi and W. Trotter, Extremal problems in discrete geometry, *Combinatorica* 3(3) (1983), 381–392.
- [32] R. Thom, Sur l’homologie des variétés algébriques réelles, in: S.S. Cairns (ed.), *Differential and Combinatorial Topology*, Princeton University Press, Princeton, NJ 1965, 255–265.

- [33] P. Ungar,  $2n$  non-collinear points determine at least  $2n$  directions, *J. Combinat. Theory* 33 (1982), 343–347.
- [34] R. Zippel, An explicit separation of relativised random polynomial time and relativised deterministic polynomial time, *Inform. Process. Lett.* 33(4) (1989), 207–212.