Unit Distances in Three Dimensions

Journal Article

Author(s):
Kaplan, Haim; Matoušek, Jiří; Safernova, Zuzana; Sharir, Micha

Publication date:
2012-07

Permanent link:
https://doi.org/10.3929/ethz-b-000049783

Originally published in:
Combinatorics, Probability & Computing 21(4), https://doi.org/10.1017/S0963548312000144

Combinatorics, Probability and Computing (2012) 21, 597–610.
doi:10.1017/S0963548312000144

Unit Distances in Three Dimensions

HAIM KAPLAN[1†], JIŘÍ MATOUŠEK[2‡],
ZUZANA SAFERNOVÁ[3‡§] and MICHA SHARIR[4¶]

[1]School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
(e-mail: haimk@tau.ac.il)
[2]Department of Applied Mathematics and Institute of Theoretical Computer Science (ITI),
Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic
(e-mail: matousek@kam.mff.cuni.cz)
and
Institute of Theoretical Computer Science, ETH Zurich, 8092 Zurich, Switzerland
[3]Department of Applied Mathematics, Charles University, Malostranské nám. 25,
118 00 Praha 1, Czech Republic
(e-mail: zuzka@kam.mff.cuni.cz)
[4]School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
and
Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA
(e-mail: michas@tau.ac.il)

Received 1 October 2011; revised 4 March 2012; first published online 25 April 2012

We show that the number of unit distances determined by n points in $\mathbb{R}^3$ is $O(n^{3/2})$, slightly improving the bound of Clarkson, Edelsbrunner, Guibas, Sharir and Welzl [5], established in 1990. The new proof uses the recently introduced polynomial partitioning technique of Guth and Katz [12]. While this paper was still in a draft stage, a similar proof of our main result was posted to the arXiv by Joshua Zahl [28].

AMS 2010 *Mathematics subject classification*: Primary 52C10

† Supported by Grant 2006/204 from the US–Israel Binational Science Foundation, and by grant 822/10 from the Israel Science Fund.
‡ Supported by the ERC Advanced Grant No. 267165.
§ Supported by the Charles University grant GAUK 421511.
¶ Supported by NSF Grant CCF-08-30272, by Grant 2006/194 from the US–Israel Binational Science Foundation, by Grant 338/09 from the Israel Science Fund, and by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University.

1. Introduction

Let P be a set of n points in Euclidean d-dimensional space $\mathbb{R}^d$. What is the maximum possible number of pairs $\{p, q\}$ of points in P such that the (Euclidean) distance of p to q is exactly 1? A standard construction, attributed to Lenz [15], shows that this number can be $\Theta(n^2)$ in $d \geqslant 4$ dimensions, so the only interesting cases are $d = 2, 3$. The planar version is the classical *unit distances* problem of Erdős [9], posed in 1946, for which we refer to the literature (in particular, see [5, 22, 24, 26]). Here we focus on the case $d = 3$. This was studied, back in 1990, by Clarkson, Edelsbrunner, Guibas, Sharir and Welzl [5], who established the upper bound $O(n^{3/2}2^{O(\alpha^2(n))})$, where $\alpha(\cdot)$ is the inverse Ackermann function (a function growing extremely slowly, much slower than $\log n$, $\log\log n$, *etc.*)

In this paper we get rid of the small factor $2^{O(\alpha^2(n))}$, and obtain the upper bound $O(n^{3/2})$. Admittedly, the improvement is not large, and achieves only a slight narrowing of the gap from the best known lower bound, which is $\Omega(n^{4/3} \log\log n)$ [10], but is nevertheless the first improvement of the bound of [5], more than 20 years after its establishment.

The proof of the new bound is based on the recently introduced *polynomial partitioning* technique of Guth and Katz [12] (see also Kaplan, Matoušek and Sharir [14] for an expository introduction). An additional goal of the present paper is to highlight certain technical issues (specifically, multi-level polynomial partitions) that might arise in the application of the new approach. These issues are relatively simple to handle for the problem at hand, but treating them in full generality is still an open issue.

Zahl's work. After we finished a draft of this paper, in early 2011, we learned that Zahl [28] had independently obtained the same bound on unit distances in $\mathbb{R}^3$ (and, actually, a more general result concerning incidences of points with suitable surfaces in $\mathbb{R}^3$), using the same general approach. Our subsequent correspondence then helped in clarifying some issues in both of the papers.

The details of our arguments differ from those of Zahl at some points, and since the general problem of the multi-level decomposition alluded to above remains unresolved (both Zahl's work and ours deal only with two-level decompositions), even slight differences in the approaches may become important in attacking the general question. Our treatment is also more pedestrian and assumes less background in algebraic geometry than Zahl's, and thus it may be more accessible for the community at large of researchers in discrete geometry. So, while we respect the priority of Zahl's arXiv preprint, and acknowledge a substantial overlap in the main ideas, we have nonetheless decided to publish our paper.

2. Analysis

Let P be a set of n points in $\mathbb{R}^3$. For each $p \in P$, let σ_p denote the unit sphere centred at p, and let Σ denote the collection of these spheres. Clearly, the number of unit distances between pairs of points of P is half the number of incidences $I(P, \Sigma)$ of the points of P with the spheres of Σ. Our main result is the following theorem.

Theorem 2.1. $I(P, \Sigma) = O(n^{3/2})$. *In particular, the number of unit distances in any set of n points in* $\mathbb{R}^3$ *is* $O(n^{3/2})$.

We first review the main algebraic ingredient of the analysis.

Polynomial partitions: A quick review. For the sake of completeness, and also for the second partitioning step in our analysis, we provide a brief review of the polynomial partitioning technique of Guth and Katz [12]; see also [14]. This technique is based on the *polynomial ham sandwich* theorem of Stone and Tukey [23]. Its specialization to three dimensions is stated in the following theorem.

Theorem 2.2 (Guth and Katz [12]). *Let P be a set of n points in $\mathbb{R}^3$ and let $s \geqslant 1$ be a parameter. Then there exists a non-zero trivariate polynomial f of degree $D = O(s^{1/3})$ and a partition of P into pairwise disjoint subsets $P_0, P_1, \ldots, P_t$, such that*

(i) $t = O(s)$,
(ii) $|P_i| \leqslant n/s$ *for each* $i = 1, \ldots, t$,
(iii) $P_0 = P \cap Z(f)$, *where* $Z(f)$ *is the zero set of* f, *and*
(iv) *each* P_i, *for* $i = 1, \ldots, t$, *is contained in a distinct connected component of* $\mathbb{R}^3 \setminus Z(f)$.

A brief review of the proof. We first recall the construction of a polynomial ham sandwich cut, as in [23], specialized to three dimensions.

We fix an integer D and put $M = \binom{D+3}{3} - 1$. Let $U_1, \ldots, U_M$ be M arbitrary finite point sets in $\mathbb{R}^3$. Let $\varphi : \mathbb{R}^3 \mapsto \mathbb{R}^M$ be the *Veronese map*, which maps a point $(x, y, z) \in \mathbb{R}^3$ to the M-tuple of the values at (x, y, z) of all the M non-constant trivariate monomials of degree at most D. We consider the images $\varphi(U_1), \ldots, \varphi(U_M)$ of our sets, and apply the standard ham sandwich theorem (see [23] and [16, Chapter 3]) to these M sets in $\mathbb{R}^M$. This yields a hyperplane h that *bisects* each set U_i, in the sense that, for each $i = 1, \ldots, M$, at most $|U_i|/2$ points of U_i lie on one side of h and at most $|U_i|/2$ points lie on the other side (the remaining points of U_i lie on h; their number can be anything between 0 and $|U_i|$).

We now consider the composition $f = h \circ \varphi$ (here $h = 0$ is the linear equation of our hyperplane). Then f is a trivariate polynomial (a linear combination of monomials) of degree at most D that bisects each of the sets $U_1, \ldots, U_M$, in the sense that, for each i,

$$|U_i \cap \{f > 0\}|,\ |U_i \cap \{f < 0\}| \leqslant |U_i|/2.$$

Note that the degree of f is at most $O(M^{1/3})$, and that its actual value can be smaller.

To prove Theorem 2.2, we apply this polynomial ham sandwich cut repeatedly, starting with the singleton set P and doubling the number of sets at each step. Specifically, we first bisect the original point set P into two halves, using a polynomial f_1. We then bisect each of these two sets into two halves, using a second polynomial f_2, bisect each of the four resulting subsets using a third polynomial f_3, and so on, until the size of all of the current subsets is reduced to at most n/s. The product $f = f_1 f_2 f_3 \cdots$ of these bisecting polynomials is the desired *partitioning polynomial*, and, as is shown in [12, 14] (and easy to verify), its degree is $D = O(s^{1/3})$.

It remains to define the subsets $P_0, P_1, \ldots, P_t$. We set $P_0 = P \cap Z(f)$; we note that we have no control over the size of P_0 – it can be anything from 0 and n. Then we let $C_1, \ldots, C_t$ be the connected components of the complement $\mathbb{R}^3 \setminus Z(f)$, and we set $P_i = P \cap C_i$, $i = 1, 2, \ldots, t$. It follows from well-known results mentioned later in Lemma 2.3

that $t = O((\deg f)^3) = O(s)$. Since each component C_i can meet at most one of the subsets produced by the sequence of the polynomial ham sandwich cuts, we have $|P_i| \leqslant n/s$ for each $i = 1, \ldots, t$. This completes the proof of the theorem. □

The intended use of the theorem is mainly with $s \leqslant n$. However, for $s > n$ we can, following the technique used in [8, 11], find a polynomial f of degree $O(n^{1/3}) = O(s^{1/3})$ that vanishes at *all* the points of P. In this case all the subsets in the resulting partition of P are empty, except for $P_0 = P \cap Z(f) = P$.

First partition. For the proof of Theorem 2.1, we set $s = n^{3/4}$, so the degree of the resulting partitioning polynomial f, yielded by Theorem 2.2, is $D = O(n^{1/4})$. Denote the resulting subsets of the above partition of P by $P_1, \ldots, P_t$, $t = O(s)$. Each of these subsets is of size at most n/s and is contained in a distinct component of $\mathbb{R}^3 \setminus Z(f)$; we also have a remainder subset P_0, contained in the zero set $Z = Z(f)$ of f.

We note that the degree D could conceivably be much smaller. For example, if P, or most of it, lies on an algebraic surface of small degree (say, a plane or a quadric) then f could be the polynomial defining that surface, resulting in a trivial partitioning in which all or most of the points of P belong to P_0 and the degree of f is very small. This potential variability of D will enter the analysis later on.

We first bound the number of incidences between $P \setminus P_0$ and Σ. For this, we need to show that no sphere crosses too many cells of the partition (that is, components of $\mathbb{R}^3 \setminus Z(f)$). This can be argued as follows.

Let us fix a sphere $\sigma = \sigma_a \in \Sigma$. The number of cells C_i crossed by σ is bounded from above by the number of components of $\sigma \setminus Z(f)$.

For bounding the latter quantity, as well as for some arguments below, it is technically convenient to use a rational parametrization of σ. Specifically, we let $\psi : \mathbb{R}^2 \to \mathbb{R}^3$ be the *inverse stereographic projection* given by $\psi(u, v) = (\psi_x(u, v), \psi_y(u, v), \psi_z(u, v))$, where

$$\psi_x(u,v) = x_0 + \frac{2u}{u^2+v^2+1}, \quad \psi_y(u,v) = y_0 + \frac{2v}{u^2+v^2+1}, \quad \psi_z(u,v) = z_0 + \frac{u^2+v^2-1}{u^2+v^2+1},$$

and (x_0, y_0, z_0) is the centre of σ. Then ψ is a homeomorphism between the uv-plane and the sphere σ 'punctured' at its north pole (recall that σ is a unit sphere). This missing point will not affect our analysis if we choose a generic coordinate frame, in which no pair of points of P are co-vertical. (Since the centre of each ball is a point in P, no point will reside at the north pole of a ball in such a generic coordinate frame.)

Let us consider the composition $f \circ \psi$ (*i.e.*, $f \circ \psi(u, v) = f(\psi_x(u, v), \psi_y(u, v), \psi_z(u, v))$); this is a rational function, which we can write as a quotient $\frac{f^*(u,v)}{q(u,v)}$ of two polynomials (with no common factor). For analysing the zero set, it suffices to consider the numerator $f^*(u, v)$, which is a polynomial of degree $O(D)$.

If f^* vanishes identically then $\sigma \subset Z(f)$ and thus σ does not cross any cell C_i of the partition. Otherwise, the number of components of $\sigma \setminus Z(f)$ is no larger than the number of components of $\mathbb{R}^2 \setminus Z(f^*)$, and for these, we use the case $d = 2$ of the following result.

Lemma 2.3. *Let f be a real polynomial of degree D in d variables. Then the number of connected components of $\mathbb{R}^d \setminus Z(f)$ is at most $6(2D)^d$.*

This lemma follows, for example, from the work of Warren [27, Theorem 2], which in turn is based on the well-known Oleinik–Petrovskii–Milnor–Thom theorem [17, 18, 25] on the sum of Betti numbers of a real algebraic variety (also see [4] for an exposition, and [1] for a neatly simplified proof of Warren's result).

From Lemma 2.3 we get that the number of connected components of $\mathbb{R}^2 \setminus Z(f^*)$ is

$$O(\deg(f^*)^2) = O(D^2).$$

We thus conclude that each sphere $\sigma = \sigma_a \in \Sigma$ crosses at most $O(D^2) = O(n^{1/2})$ cells C_i of the partition.

Hence the overall number of sphere–cell crossings is $O(nD^2) = O(n^{3/2})$. Now we can estimate the number of incidences of the spheres with the points of $P_1, \ldots, P_t$ in the following standard manner. For $i = 1, \ldots, t$, let $P_i = P \cap C_i$ be the set of points inside a cell C_i, and let n_i be the number of spheres crossing C_i. Then the spheres crossing C_i and incident to at most two points of P_i contribute at most $2n_i$ incidences, which, summed over all C_i, amounts to at most $O(n^{3/2})$ incidences. It remains to deal with spheres incident to at least three points of P_i, and here we observe that for a fixed point $p \in P_i$, the number of spheres that are incident to p and contain at least two other points of P_i is at most $2\binom{|P_i|-1}{2} \leqslant |P_i|^2$, because any pair of points $q, r \in P_i \setminus \{p\}$ determine at most two unit spheres that are incident to p, q, r. Hence the number of incidences of the points of P_i with spheres that are incident to at least three points of P_i is at most $|P_i|^3 \leqslant (n/s)^3 = O(n^{3/4})$. Summing over all subsets P_i, we get a total of $O(n^{3/2})$ such incidences.[1]

Remark. (Although the full significance of this remark will become clearer later on, we nevertheless make it early in the game.) There are well-known papers in real algebraic geometry estimating the number of components of algebraic varieties in $\mathbb{R}^d$, or more generally, the complexity of an arrangement of zero sets of polynomials in $\mathbb{R}^d$ (Oleinik and Petrovskiĭ [18], Milnor [17], Thom [25], and Warren [27]). In the arguments used so far and also below, we need bounds in a somewhat different setting, namely, when the arrangement is not in $\mathbb{R}^d$, but within some algebraic variety. This setting was considered by Basu, Pollack and Roy [3]. However, their bound is not sufficiently sharp for us either, since it assumes the same upper bound both on the degree of the polynomials defining the arrangement and those defining the variety. Prompted by our question, Barone and Basu [2] recently proved a bound in this setting involving two degree parameters: they consider a k-dimensional variety V in $\mathbb{R}^d$ defined by polynomials of degree at most D, and an arrangement of n zero sets of polynomials of degree at most E within V, and they bound the number of cells, of all dimensions, in the arrangement by $O(1)^d D^{d-k}(nE)^k$. A weaker bound of a similar kind was also derived independently by Solymosi and Tao [21, Theorem A.2].

We could refer to the Barone–Basu result in the proof above, instead of using the rational parametrization and Lemma 2.3. However, later on we will need three different

[1] Alternatively, we can use the Kővári–Sós–Turán theorem (see [19]) on the maximum number of edges in a bipartite graph with a forbidden $K_{r,s}$ subgraph, as was done in many previous papers; this comment applies to several similar arguments below.

degree parameters (involving spheres intersecting a variety defined by two polynomials of two potentially different degrees; in this case one of the degrees is 2, the degree of the polynomial equation of a sphere), and thus we cannot refer to [2, 21] directly. We provide elementary *ad hoc* arguments instead (aimed mainly at readers not familiar with the techniques employed in [2, 21]). If the multi-level polynomial partition method should be used in dimensions higher than 3, a more systematic approach will be needed to bound the appropriate number of components. We believe that the approach of [2] should generalize to an arbitrary number of different degree parameters, but there are several other obstacles to be overcome along the way; see Section 3 for a discussion. This is the end of the longish remark, and we come back to the proof.

Bounding $I(P_0,\Sigma)$. It therefore remains to bound $I(P_0,\Sigma)$. Here is an informal overview of this second step of the analysis. We apply the polynomial partitioning procedure to P_0, using a second polynomial g (which again is the product of logarithmically many bisecting polynomials). For a good choice of g, we will obtain various subsets of P_0 of roughly equal sizes, lying in distinct components of $Z(f)\setminus Z(g)$, and a remainder subset $P_{00}\subset Z(f)\cap Z(g)$. Again, for a good choice of g, $Z(f)\cap Z(g)$ will be a one-dimensional curve, and it will be reasonably easy to bound $I(P_{00},\Sigma)$. The situation that we want to avoid is one in which f and g have a common factor, whose two-dimensional zero set contains most of P_0, in which case the dimension reduction that we are after (from a two-dimensional surface to a one-dimensional curve) will not work.

To overcome this potential problem, we first factor f into irreducible factors $f = f_1 f_2\cdots f_r$ (recall that in the construction of [12], f is the product of logarithmically many factors, some of which may themselves be reducible). Denote the degree of f_i by D_i, so $\sum_i D_i = D$. By removing repeated factors from f, if any exist, we may assume that f is square-free; this does not affect the partition induced by f, nor its zero set. Put

$$\begin{aligned} P_{01} &= P_0\cap Z(f_1)\\ P_{02} &= \bigl(P_0\setminus P_{01}\bigr)\cap Z(f_2)\\ &\cdots\\ P_{0i} &= \Bigl(P_0\Bigm\backslash\bigcup_{j<i}P_{0j}\Bigr)\cap Z(f_i)\\ &\cdots \end{aligned}$$

This is a partition of P_0 into r pairwise disjoint subsets. Put $m_i = |P_{0i}|$ for $i=1,\ldots,r$; thus, $\sum_i m_i \leqslant n$. We will bound $I(P_{0i},\Sigma)$ for each i separately and then add up the resulting bounds to get the desired bound on $I(P_0,\Sigma)$.

Second partition. We will bound the number of incidences between P_{0i} and Σ using the following lemma, which is the core of (this step of) our analysis.

Lemma 2.4. *Let f be an irreducible trivariate polynomial of degree D, let Q be a set of m points contained in $Z(f)$, and let Σ be a set of $n\geqslant m$ unit spheres in $\mathbb{R}^3$. Then*

$$I(Q,\Sigma) = O\bigl(m^{3/5}n^{4/5}D^{2/5} + nD^2\bigr).$$

Remark. When $D = 1$ (all the points of Q are co-planar), the bound in the lemma becomes $O(m^{3/5}n^{4/5} + n)$, a special case (when $m \leqslant n$) of the bound $O(m^{3/5}n^{4/5} + n + m)$, which is a well-known upper bound on the number of incidences between m points and n circles in the plane (see, *e.g.*, [5, 20]). In our case, the circles are the intersections of the spheres of Σ with the plane (where each resulting circle has multiplicity at most 2).

The main technical step in proving Lemma 2.4 is encapsulated in the following lemma.

Lemma 2.5. *Given an irreducible trivariate polynomial f of degree D, a parameter $E \geqslant D$, and a finite point set Q in $\mathbb{R}^3$, there is a polynomial g of degree at most E, co-prime with f, which partitions Q into subsets $Q_0 \subseteq Z(g)$ and $Q_1, \ldots, Q_t$, for $t = \Theta(DE^2)$, so that each Q_i, for $i = 1, \ldots, t$, lies in a distinct component of $\mathbb{R}^3 \setminus Z(g)$, and $|Q_i| = O(|Q|/t)$.*

Note the similarity of this lemma to the standard polynomial partitioning result (Theorem 2.2). The difference is that, to ensure that g be co-prime with f, we pay the price of having only $\Theta(DE^2)$ parts in the resulting partition, instead of $\Theta(E^3)$.

Proof of Lemma 2.5. As in the standard polynomial partitioning technique, we obtain g as the product of logarithmically many bisecting polynomials, each obtained by applying a variant of the polynomial ham sandwich theorem to a current collection of subsets of Q. The difference, though, is that we want to ensure that each of the bisecting polynomials is not divisible by f; since f is irreducible, this ensures co-primality of g with f. Reviewing the construction of polynomial ham sandwich cuts, as outlined in the proof of Theorem 2.2, we see that all that is needed is to come up with some sufficiently large finite set of monomials, of an appropriate maximum degree, so that no non-trivial linear combination of these monomials can be divisible by f. We then use a restriction of the Veronese map defined by this subset of monomials, and the standard ham sandwich theorem in the resulting high-dimensional space, to obtain the desired polynomial.

Let $x^i y^j z^k$ be the *leading term* of f, in the sense that $i + j + k = D$ and (i, j, k) is largest in the lexicographical order among all the triples of exponents of the monomials of f (with non-zero coefficients) of degree D. Let q be the desired number of sets that we want a single partitioning polynomial to bisect. For that we need a space of q monomials whose degrees are not too large and which span only polynomials not divisible by f. If, say, $q < \left(\frac{D}{3}\right)^3$ then we can use all monomials $x^{i'} y^{j'} z^{k'}$ such that $i', j', k' \leqslant q^{1/3} < D/3$. Clearly, any non-trivial linear combination of these monomials cannot be divisible by f. In this case the degree of the resulting partitioning polynomial is $\Theta(q^{1/3})$. If $q \geqslant \left(\frac{D}{3}\right)^3$ then we take the set of all monomials $x^{i'} y^{j'} z^{k'}$ that satisfy $i' < i$ or $j' < j$ or $k' < k$, and $\max\{i', j', k'\} \leqslant \tilde{D}$ for a suitable integer $\tilde{D}$, which we specify below (the actual degree of the bisecting polynomial under construction will then be at most $3\tilde{D}$). Any non-trivial polynomial h which is a linear combination of these monomials cannot be divisible by f. Indeed, if $h = fh_1$ for some polynomial h_1 then the product of the leading terms of f and of h_1 cannot be cancelled out by the other monomials of the product, and, by construction, h cannot contain this monomial. The number of monomials in this set is $\Theta(i\tilde{D}^2 + j\tilde{D}^2 + k\tilde{D}^2) = \Theta(D\tilde{D}^2)$. We thus pick $\tilde{D} = \Theta((q/D)^{1/2})$ so that we indeed get q monomials. As noted above, the degree of the resulting bisecting polynomial in this case is $O((q/D)^{1/2})$.

We now proceed to construct the required partitioning of Q into t sets, by a sequence of about $\log t$ polynomials $g_0, g_1, \ldots$, where g_j bisects 2^j subsets of Q, each of size at most $|Q|/2^j$. For every j such that $q = 2^j < \left(\frac{D}{3}\right)^3$ we construct, as shown above, a polynomial of degree $O(q^{1/3}) = O(2^{j/3})$. For the indices j with $q = 2^j \geqslant \left(\frac{D}{3}\right)^3$, we construct a polynomial of degree $O(\left(q/D\right)^{1/2}) = O(2^{j/2}/D^{1/2})$. Since the upper bounds on the degrees of the partitioning polynomials increase exponentially with j, and since the number of parts t that we want is $\Omega(D^3)$ (we want it to be $\Theta(DE^2)$ and $E \geqslant D$), it follows that the degree of the product of the sequence is $O(\left(t/D\right)^{1/2})$. If we require this degree bound to be no larger than E, then it follows that the size of the partition that we get is $t = \Theta(DE^2)$. Clearly, f does not divide the product g of the polynomials g_j, so g satisfies all the properties asserted in the lemma. □

Remarks. (1) The analysis given above can be interpreted as being applied to the *quotient ring* $\mathbb{R}[x, y, z]/I$, where $I = \langle f \rangle$ is the ideal generated by f. General quotient rings are described in detail in, *e.g*., [6, 7], but the special case where I is generated by a single polynomial is much simpler, and can be handled in the simple manner described above, bypassing (or rather simplifying considerably) the general machinery of quotient rings. As a matter of fact, an appropriate extension of Lemma 2.5 to quotient rings defined by two or more polynomials is still an open issue; see Section 3.

(2) The set Q is in fact contained in $Z(f)$, and the subset Q_0 is contained in $Z(f) \cap Z(g)$. However, except for the effect of this property on the specific choice of monomials for g, the construction considers Q as an arbitrary set of points in $\mathbb{R}^3$, and does not exploit the fact that $Q \subset Z(f)$.

Back to the proof of Lemma 2.4. We apply Lemma 2.5 to Q, now assumed to be contained in $Z(f)$, and obtain the desired partitioning polynomial g. We now proceed, based on the resulting partition of Q, to bound $I(Q, \Sigma)$; we follow the notation used in Lemma 2.5.

We need the following technical lemma, a variant of which has been established and exploited in [11] and in [8]. For the sake of completeness we include a brief sketch of its proof, and refer the reader to the aforementioned papers for further details.

Lemma 2.6.

(a) *Let f and g be two trivariate polynomials of respective degrees D and E. Let Π be an infinite collection of parallel planes such that, for each $\pi \in \Pi$, the restrictions of f and g to Π have more than DE common roots. Then f and g have a* (*non-constant*) *common factor.*

(b) *Let f and g be as in* (a). *If the intersection $Z(f) \cap Z(g)$ of their zero sets contains a two-dimensional surface patch then f and g have a* (*non-constant*) *common factor.*

Sketch of proof. **(a)** Assume without loss of generality that the planes in Π are horizontal and that, if the number of common roots in a plane is finite then these roots have different x-coordinates; both assumptions can be enforced by an appropriate rotation of the coordinate frame. Consider the y-resultant $r(x,z) = \mathrm{Res}_y(f(x,y,z), g(x,y,z))$ of $f(x,y,z)$ and $g(x,y,z)$. This is a polynomial in x and z of degree at most DE. If the plane $z = c$ contains more than DE common roots then $r(x,c)$, which is a polynomial in x, has more than DE roots, and therefore it must be identically zero. It follows that $r(x,z)$ is identically zero on infinitely many planes $z = c$, and therefore it must be identically zero. (Its restriction to an arbitrary non-horizontal line ℓ has infinitely many roots and therefore it must be identically zero on ℓ.) It follows that $f(x,y,z)$ and $g(x,y,z)$ have a common factor (see [6, Proposition 1, page 163]).

(b) This follows from (a), since if $Z(f)$ and $Z(g)$ contain a two-dimensional surface patch, then they must have infinitely many zeros on infinitely many parallel planes. □

Incidences outside $Z(g)$. To prove Lemma 2.4, we first bound the number of incidences of the points of a fixed subset Q_j, for $j \geqslant 1$, with Σ, using the same approach as in the first partition. That is, let n_j denote the number of spheres of Σ that cross the corresponding cell C_j *effectively*, in the sense that $\sigma \cap Q_j \neq \emptyset$. Then we have $O(n_j)$ incidences of the points of Q_j with spheres that are incident to at most two points of Q_j, and $O((m/t)^3)$ incidences with spheres that are incident to at least three points. Summing over all sets, we get

$$\sum_{j=1}^{t} I(Q_j, \Sigma) = O\left(m^3/t^2 + \sum_{j=1}^{t} n_j\right). \tag{2.1}$$

We estimate $\sum_j n_j$ by bounding the number of cells C_j that a single sphere $\sigma \in \Sigma$ can cross effectively, which we do as follows.

Take the same rational parametrization ψ of σ used in the analysis of the first partitioning step. Let $f^*(u,v)$ and $g^*(u,v)$ be the polynomials obtained from $f \circ \psi$ and $g \circ \psi$ by removing the common denominators of these rational functions. The degrees of f^* and g^* are $O(D)$ and $O(E)$, respectively.

If f^* vanishes identically on the uv-plane, then $\sigma \subseteq Z(f)$; this is an easy situation that we will handle later on. Otherwise, $Z(f^*) = \psi^{-1}(\sigma \cap Z(f))$ is a one-dimensional curve γ in the uv-plane (possibly degenerate, *e.g.*, empty or consisting of isolated points), and $Q \cap \sigma$ is contained in $\psi(\gamma)$.

By construction, the number of cells C_j that σ crosses effectively (so that it is incident to points of Q_j) is no larger than the number of components of $Z(f^*) \setminus Z(g^*)$. This is because each such cell C_j contains at least one connected component of $\psi(Z(f^*) \setminus Z(g^*))$.

Now each component of $Z(f^*) \setminus Z(g^*)$ is either a full component of $Z(f^*)$, or a relatively open connected portion of $Z(f^*)$ whose closure meets $Z(g^*)$.

Since f^* is a bivariate polynomial, Harnack's theorem [13] asserts that the number of (arcwise) connected components of $Z(f^*)$ is at most $1 + \binom{\deg(f^*)-1}{2} = O(D^2)$.

For the other kind of components, choose a generic sufficiently small value $\varepsilon > 0$, so that f^* and $g^* \pm \varepsilon$ do not have a common factor.[2] Then each component of $Z(f^*) \setminus Z(g^*)$ of the second kind must contain a point at which $g^* + \varepsilon = 0$ or $g^* - \varepsilon = 0$. Hence, the number of such components is at most the number of such common roots, which, by Bézout's theorem (see, *e.g.*, [7]) is[3] $O(\deg(f^*)\deg(g^*)) = O(DE)$.

Since $E \geqslant D$, we conclude that the number of cells C_j crossed effectively by σ is $O(DE)$, which in turn implies that $\sum_j n_j = O(nDE)$. Substituting this in (2.1) and recalling that $t = \Theta(DE^2)$, we get

$$\sum_{j=1}^{r} I(Q_j, \Sigma) = O\left(\frac{m^3}{D^2E^4} + nDE\right). \tag{2.2}$$

We have left aside the case where $\sigma \subseteq Z(f)$. Since f is irreducible, and so is σ, we must have $\sigma = Z(f)$ in this case (recall Lemma 2.6(b)). The analysis proceeds as above for every sphere $\sigma' \neq \sigma$, and the number of incidences with σ itself is at most m, a bound that is subsumed by the bound asserted in the lemma (recall that $m \leqslant n$).

We note that in the ongoing analysis D is the actual degree of the irreducible factor of f under consideration, but E is only a chosen upper bound for $\deg(g)$, whose actual value may be smaller (as may have been the case with f).

To optimize the bound in (2.2), we choose

$$E = \max\left\{\frac{m^{3/5}}{n^{1/5}D^{3/5}},\ D\right\}, \tag{2.3}$$

and observe that the first term dominates when $D \leqslant m^{3/8}/n^{1/8}$. Assuming that this is indeed the case, we get

$$\sum_j I(Q_j, \Sigma) = O(m^{3/5}n^{4/5}D^{2/5}). \tag{2.4}$$

If $D > m^{3/8}/n^{1/8}$ then we have $E = D$, and the bound (2.2) becomes

$$\sum_j I(Q_j, \Sigma) = O\left(\frac{m^3}{D^6} + nD^2\right) = O(nD^2). \tag{2.5}$$

Thus, $I(Q \setminus Q_0, \Sigma)$ satisfies the bound asserted in the lemma, and it remains to bound $I(Q_0, \Sigma)$.

Incidences within $Z(f) \cap Z(g)$. Recall that Q_0 is contained in the curve $\delta = Z(f) \cap Z(g)$, which by Lemma 2.6(b) is (at most) one-dimensional.

Fix a sphere $\sigma \in \Sigma$ that does not coincide with $Z(f)$, let ψ be the corresponding rational parametrization of σ, and let f^*_σ and g^*_σ be the numerators of $f \circ \psi$ and $g \circ \psi$, as defined in the preceding analysis.

[2] Indeed, assuming that f^* and $g^* + \varepsilon$ had a non-constant common factor for infinitely many values of ε, then the same factor would occur for two distinct values ε_1 and ε_2 of ε, and thus it would have to divide $\varepsilon_1 - \varepsilon_2$, which is impossible.

[3] The $O(DE)$ bound for the number of components of $Z(f^*) \setminus Z(g^*)$ is also a direct consequence of the main result of Barone and Basu [2].

If g^*_σ is identically 0, then we have $\sigma \subseteq Z(g)$, and the irreducible quadratic polynomial defining σ is a factor of g by Lemma 2.6. Thus, the number of such σ is $O(E)$, and together they can contribute at most $O(mE)$ incidences, which is bounded from above by the right-hand side of (2.2). The case of $f^*_\sigma \equiv 0$ has been assumed not to occur.

We therefore assume that both f^*_σ and g^*_σ are not identically zero, we let h^*_σ denote the greatest common divisor of f^*_σ and g^*_σ, and put $f^*_\sigma = f^*_{1\sigma} h^*_\sigma$ and $g^*_\sigma = g^*_{1\sigma} h^*_\sigma$. Then $\psi^{-1}(\sigma \cap \delta)$ is the union of $Z(h^*_\sigma)$ and of $Z(f^*_{1\sigma}) \cap Z(g^*_{1\sigma})$. Using Bézout's theorem as above, we have $|Z(f^*_{1\sigma}) \cap Z(g^*_{1\sigma})| = O(DE)$; summing this bound over all spheres σ, we get at most $O(nDE)$ incidences, a bound already subsumed by (2.2).

It remains to account for incidences of the following kind (call them h^*-*incidences*): a point $q \in Q_0 \cap \sigma$ lying in $\psi(Z(h^*_\sigma))$. Let us call such a point q *isolated in* σ if it is an isolated point of $\psi(Z(h^*_\sigma))$; *i.e.*, there is a neighbourhood of q in σ intersecting $\psi(Z(h^*_\sigma))$ only at q.

The homeomorphism ψ^{-1} maps the isolated points q on σ to isolated zeros of h^*_σ in the uv-plane, in a one-to-one fashion. Since $\deg(h^*_\sigma) = O(D)$, $Z(h^*_\sigma)$ has at most $O(D^2)$ components (Harnack's theorem again), and thus the overall number of isolated incidences is $O(nD^2)$.

Finally, to account for non-isolated h^*-incidences, let us fix a point $q \in Q_0$, and consider the collection $\tilde{\Sigma}_q$ consisting of all spheres $\sigma \in \Sigma$ that contain q such that q forms a non-isolated h^*-incidence with σ. We claim that $|\tilde{\Sigma}_q| = O(DE)$.

For $\sigma \in \tilde{\Sigma}_q$, the set $\psi(Z(h^*_\sigma))$ contains a curve segment $\beta_{q,\sigma}$ ending at q. Let us call $\beta_{q,\sigma}$ and $\beta_{q,\sigma'}$ *equivalent* if they coincide in some neighbourhood of q. If $\beta_{q,\sigma}$ and $\beta_{q,\sigma'}$ are not equivalent, then in a sufficiently small neighbourhood of q they intersect only at q (since they are arcs of algebraic curves).

We also note that a given $\beta_{q,\sigma}$ can be equivalent to $\beta_{q,\sigma'}$ for at most one $\sigma' \neq \sigma$; this is because the common portion $\beta_{q,\sigma} \cap \beta_{q,\sigma'}$ of the considered curve segments has to be contained in the intersection circle $\sigma \cap \sigma'$, and that circle intersects any other sphere $\sigma'' \in \Sigma$ in at most two points. Thus, $|\tilde{\Sigma}_q|$ is at most twice the number of equivalence classes of the curve segments $\beta_{q,\sigma}$.

Let us fix an auxiliary sphere S of a sufficiently small radius ρ around q, so that each $\beta_{q,\sigma}$ intersects S at some point x_σ. Let S' be a sphere around q of radius $\varepsilon\rho$, for some sufficiently small constant parameter $\varepsilon > 0$. We choose a point $y \in S'$ uniformly at random, and let π be the plane tangent to S' at y. Then, for each $\sigma \in \tilde{\Sigma}_q$, π separates x_σ from q with probability at least $\frac{1}{3}$, say (which can be guaranteed by choosing ε sufficiently small), and thus, by continuity, it intersects $\beta_{q,\sigma}$. Hence there is a specific $y_0 \in S'$ such that the corresponding tangent plane π_0 intersects $\beta_{q,\sigma}$ for at least a third of the spheres $\sigma \in \tilde{\Sigma}_q$.

Moreover, we can assume that such a π_0 intersects each $\beta_{q,\sigma}$ in such a way that all planes π parallel to π_0 and sufficiently close to it intersect $\beta_{q,\sigma}$ as well. Then an application of Lemma 2.6(a) allows us to assume that the restrictions of f and g to some π as above (actually to most of these planes) are bivariate polynomials, with at most DE common roots. Hence π intersects at most $O(DE)$ of the curves $\beta_{q,\sigma}$, and so $|\tilde{\Sigma}_q| = O(DE)$.

Altogether, we can bound the number of h^*-incidences by $O(nD^2 + mDE)$, which does not exceed the earlier estimate $O(nDE)$ (recalling that $m \leqslant n$). Hence, choosing E as in (2.3), the incidences within δ do not affect either of the asymptotic bounds (2.4) and (2.5).

This completes the proof of Lemma 2.4. □

Finishing the proof of Theorem 2.1. We recall that in the first partitioning step, the set $P_0 = P \cap Z(f)$ was partitioned into the subsets $P_{01}, \ldots, P_{0r}$. Each P_{0i} consists of m_i points and is contained in $Z(f_i)$, where f_i is an irreducible factor of f, with $\deg(f_i) = D_i$. By Lemma 2.4 we have

$$\sum_{i=1}^{r} I(P_{0i}, \Sigma) = O\left(\sum_{i=1}^{r} m_i^{3/5} n^{4/5} D_i^{2/5} + \sum_{i=1}^{r} n D_i^2\right).$$

For the first term on the right-hand side we use Hölder's inequality[4] and the inequalities

$$\sum_{i=1}^{r} D_i \leqslant D = O(n^{1/4}) \quad \text{and} \quad \sum_{i=1}^{r} m_i \leqslant n.$$

Thus,

$$n^{4/5} \sum_{i=1}^{r} m_i^{3/5} D_i^{2/5} \leqslant n^{4/5} \left(\sum_i m_i\right)^{3/5} \left(\sum_i D_i\right)^{2/5} \leqslant O(n^{4/5} n^{3/5} D^{2/5}) = O(n^{3/2}).$$

For the remaining term we have

$$\sum_{i=1}^{r} n D_i^2 \leqslant nD \cdot \sum_{i=1}^{r} D_i \leqslant nD^2 = O(n^{3/2}).$$

We thus get a total of $O(n^{3/2})$ incidences, thereby completing the proof of the theorem. □

3. Discussion

The main technical ingredient in the analysis, on top of the standard polynomial partitioning technique of Guth and Katz, is the recursion on the dimension of the ambient manifold containing the points of P. This required a more careful construction of the second partitioning polynomial g to make sure that it is co-prime with the first polynomial f. It is reasonably easy to perform the first such recursive step, as done here and also independently by Zahl [28], but successive recursive steps become trickier. In such cases we have several co-prime polynomials, and we need to construct, in the quotient ring of their ideal, a polynomial ham sandwich cut of some specified maximum degree with sufficiently many monomials. Such higher recursive steps will be needed when we analyse incidences between points and surfaces in higher dimensions. At the moment there does not seem to be an efficient procedure for this task. Another recent paper where similar issues arise is by Solymosi and Tao [21].

We also note that Zahl's study extends Theorem 2.1 to incidences between points and more general surfaces in three dimensions. The analysis in our study can also be similarly

[4] Hölder's inequality asserts that $\sum x_i y_i \leqslant (\sum |x_i|^p)^{1/p} (\sum |y_i|^q)^{1/q}$ for positive p, q satisfying $1/p + 1/q = 1$. Here we use it with $p = 5/3$, $q = 5/2$, $x_i = m_i^{3/5}$, and $y_i = D_i^{2/5}$.

extended (at the price of making some of the arguments more complicated), but, since our goal had been to improve the bound on unit distances, we have focused on the case of unit spheres.

Acknowledgement

We would like to thank Josh Zahl for an e-mail discussion and useful comments.

References

[1] Akama, Y., Irie, K., Kawamura, A. and Uwano, Y. (2010) VC dimensions of principal component analysis. *Discrete Comput. Geom.* **44** 589–598.

[2] Barone, S. and Basu, S. (2012) Refined bounds on the number of connected components of sign conditions on a variety. *Discrete Comput. Geom.* **47** (3) 577–597.

[3] Basu, S., Pollack, R. and Roy, M.-F. (1996) On the number of cells defined by a family of polynomials on a variety. *Mathematika* **43** 120–126.

[4] Basu, S., Pollack, R. and Roy, M.-F. (2003) *Algorithms in Real Algebraic Geometry*, Vol. 10 of *Algorithms and Computation in Mathematics*, Springer.

[5] Clarkson, K., Edelsbrunner, H., Guibas, L., Sharir, M. and Welzl, E. (1990) Combinatorial complexity bounds for arrangements of curves and spheres. *Discrete Comput. Geom.* **5** 99–160.

[6] Cox, D., Little, J. and O'Shea, D. (2005) *Using Algebraic Geometry*, second edition, Springer.

[7] Cox, D., Little, J. and O'Shea, D. (2007) *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, third edition, Springer.

[8] Elekes, G., Kaplan, H. and Sharir, M. (2011) On lines, joints, and incidences in three dimensions. *J. Combin. Theory Ser. A* **118** 962–977. Also in `arXiv:0905.1583`.

[9] Erdős, P. (1946) On a set of distances of n points. *Amer. Math. Monthly* **53** 248–250.

[10] Erdős, P. (1960) On sets of distances on n points in Euclidean space. *Magyar Tud. Akad. Mat. Kutató Int. Kozl.* **5** 165–169.

[11] Guth, L. and Katz, N. H. (2010) Algebraic methods in discrete analogs of the Kakeya problem, *Adv. Math.* **225** 2828–2839. Also in `arXiv:0812.1043v1`.

[12] Guth, L. and Katz, N. H. (2010) On the Erdős distinct distances problem in the plane. `arXiv:1011.4105`.

[13] Harnack, C. G. A. (1876) Über die Vielfaltigkeit der ebenen algebraischen Kurven. *Math. Ann.* **10** 189–199.

[14] Kaplan, H., Matoušek, J. and Sharir, M. (2011) Simple proofs of classical theorems in discrete geometry via the Guth–Katz polynomial partitioning technique. *Discrete Comput. Geom.*, submitted. Also in `arXiv:1102.5391`.

[15] Lenz, H. (1955) Zur Zerlegung von Punktmengen in solche kleineren Durchmessers. *Arch. Math.* **6** 413–416.

[16] Matoušek, J. (2003) *Using the Borsuk–Ulam Theorem*, Springer.

[17] Milnor, J. (1964) On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.* **15** 275–280.

[18] Oleinik, O. A. and Petrovskii, I. B. (1949) On the topology of real algebraic surfaces. *Izv. Akad. Nauk SSSR* **13** 389–402.

[19] Pach, J. and Agarwal, P. K. (1995) *Combinatorial Geometry*, Wiley-Interscience.

[20] Pach, J. and Sharir, M. (2004) Geometric incidences. In *Towards a Theory of Geometric Graphs* (J. Pach, ed.), Vol. 342 of *Contemporary Mathematics*, AMS, pp. 185–223.

[21] Solymosi, J. and Tao, T. (2012) An incidence theorem in higher dimensions. `arXiv:1103.2926v4`

[22] Spencer, J., Szemerédi, E. and Trotter, W. T. (1984) Unit distances in the Euclidean plane. In *Graph Theory and Combinatorics: Proc. Cambridge Conf. on Combinatorics* (B. Bollobás, ed.), Academic Press, pp. 293–308.
[23] Stone, A. H. and Tukey, J. W. (1942) Generalized sandwich theorems. *Duke Math. J.* **9** 356–359.
[24] Székely, L. (1997) Crossing numbers and hard Erdős problems in discrete geometry. *Combinat. Probab. Comput.* **6** 353–358.
[25] Thom, R. (1965) Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology* (S. S. Cairns, ed.), Princeton University Press, pp. 255–265.
[26] Valtr, P. (2006) Strictly convex norms allowing many unit distances and related touching questions, manuscript.
[27] Warren, H. E. (1968) Lower bound for approximation by nonlinear manifolds. *Trans. Amer. Math. Soc.* **133** 167–178.
[28] Zahl, J. (2011) An improved bound on the number of point-surface incidences in three dimensions. `arXiv:1104.4987`. First posted (v1) 26 April 2011; revised and corrected 22 September 2011.