JOINTS OF VARIETIES

JONATHAN TIDOR, HUNG-HSUN HANS YU, AND YUFEI ZHAO

Abstract. We generalize the Guth–Katz joints theorem from lines to varieties. A special case says that N planes (2-flats) in 6 dimensions (over any field) have $O(N^{3/2})$ joints, where a joint is a point contained in a triple of these planes not all lying in some hyperplane. More generally, we prove the same bound when the set of N planes is replaced by a set of 2-dimensional algebraic varieties of total degree N, and a joint is a point that is regular for three varieties whose tangent planes at that point are not all contained in some hyperplane. Our most general result gives upper bounds, tight up to constant factors, for joints with multiplicities for several sets of varieties of arbitrary dimensions (known as Carbery's conjecture). Our main innovation is a new way to extend the polynomial method to higher dimensional objects, relating the degree of a polynomial and its orders of vanishing on a given set of points on a variety.

1. Introduction

Guth and Katz [18] proved the following "joints theorem": N lines in $\mathbb{R}^3$ have $O(N^{3/2})$ joints, where a *joint* is a point contained in three of the lines that do not all lie on some plane. This bound is tight up to a constant factor due to the following example: consider k generic planes—their pairwise intersections give $\binom{k}{2}$ lines and triplewise intersections give $\binom{k}{3}$ joints.

The joints problem was first studied in Chazelle et al. [7]. Besides being an interesting problem in incidence geometry, it also caught the attention of harmonic analysts due to connections to the Kakeya problem as observed by Wolff [34]. This connection was further elucidated by Bennett, Carbery and Tao [1] in their work on the multilinear Kakeya problem, which in turn allowed them to improve bounds on the joints problem (prior to the Guth–Katz solution). Guth [13] later adapted techniques from the solution of the joints theorem to prove the so-called endpoint case of the Bennett–Carbery–Tao multilinear Kakeya conjecture, which can be viewed as a joints theorem for tubes (also see the exposition in [15, Section 15.8]). Guth's multilinear Kakeya result was later generalized by Zhang [38] to slabs and neighborhoods of varieties (though the latter does not translate back to the joints problem for flats).

The Guth–Katz solution of the joints problem highlights the importance of the polynomial method. Their joints theorem was also a precursor to their subsequent breakthrough on the Erdős distinct distances problem [19], which introduced a polynomial partitioning method that has found many subsequent applications. One of the key steps in [19] dealt with a point-line incidence problem in $\mathbb{R}^3$ with additional constraints on the configuration of lines. These developments were partly inspired by Dvir's [8] stunningly short and elegant solution to the finite field Kakeya problem. Guth has also successfully applied the polynomial method developed in this line of work to restriction problems related to Kakeya [16, 17].

Since Guth and Katz's original work, there has been significant effort in extending the joints theorem [4, 5, 6, 12, 20, 21, 22, 23, 24, 25, 26, 37, 39]. Kaplan, Sharir, and Shustin [25] and Quilodrán [26] independently extended the joints theorem from $\mathbb{R}^3$ to $\mathbb{R}^d$, and these techniques and results extend to arbitrary fields as stated below (also see [4, 9, 29]). Given a set of lines in $\mathbb{F}^d$, a *joint* is a point contained in d lines with independent and spanning directions. Throughout the paper, $\mathbb{F}$ stands for an arbitrary field, and our constants do not depend on $\mathbb{F}$.

Theorem 1.1. *A set of N lines in $\mathbb{F}^d$ has at most $C_d N^{d/(d-1)}$ joints, for some constant C_d.*

Recently Yu and Zhao [37] proved that N lines in $\mathbb{F}^d$ have at most $\frac{(d-1)!^{1/(d-1)}}{d} N^{d/(d-1)}$ joints. This leading constant is optimal, matching the above construction up to a $(1+o(1))$-factor.

We generalize the joints theorem from lines to varieties, overcoming a fundamental difficulty with the polynomial method that one quickly runs into—we will elaborate more on this later. A representative case of our result says the following. Here a *joint* is a point contained in a triple of planes not all lying in some hyperplane. All our bounds on joints in this paper are tight up to a constant factor (depending on the dimension) due to a straightforward generalization of the example in the first paragraph.

Theorem 1.2. *A set of N planes in $\mathbb{F}^6$ has $O(N^{3/2})$ joints.*

In his PhD thesis, Ben Yang [35, 36] proved partial results giving an upper bound $N^{3/2+o(1)}$ when $\mathbb{F} = \mathbb{R}$ (and also more generally for bounded degree varieties in $\mathbb{R}^d$—in contrast, our results on joints of varieties do not require any bounded degree hypotheses). Yang's results have two fundamental limitations: (1) an error term in the exponent and (2) the methods only work over the reals. He used a variant of the polynomial partitioning method [19], which requires real topology. More specifically, Yang applied polynomial partitioning for varieties (due to Guth [14] and extended by Blagojević, Blagojević, and Ziegler [2]) using bounded degree polynomials (due to Solymosi and Tao [27]), with the latter requiring an error term in the exponent. We introduce a novel approach that avoids both limitations.

The only other prior result on joints of higher dimensional objects says that, as a representative example, a set of L lines and F planes in $\mathbb{F}^4$ has $O(LF^{1/2})$ joints, where now a joint is defined to be a point contained in two lines and one plane, not all lying on a hyperplane (this result was recently independently proved by Yu and Zhao [37] and Carbery and Iliopoulou [5]; Yang mentioned at the end of his thesis [36] that he could also obtain this claim, though without details). Even the "next" case of "line-plane-plane" joints was open before this work.

Incidence geometry and the polynomial method concerning higher dimensional objects often tend to be substantially more intricate compared to problems that only involve lines and points. Our work introduces a new way to tackle such problems. Let us highlight some other representative works on higher dimensional incidence problems. Solymosi and Tao [27] introduced a bounded degree variation of the polynomial partitioning method, used in Yang's proof mentioned earlier, to give nearly tight (up to a $+o(1)$ error term in the exponent) bound for incidences between points and k-dimensional varieties of bounded degree in $\mathbb{R}^d$, in the spirit of the Szemerédi–Trotter theorem [28] for point-line incidences in the plane. Using different methods, Walsh [32, 33] recently developed powerful techniques for understanding incidences between sets of m-dimensional and $m+1$-dimensional varieties, thereby unifying a large body of incidence geometry results in the literature. However, we do not see how to apply Walsh's techniques for extending the joints theorem. The above approaches use different forms of "partitioning" and involve iteratively restricting the ambient space to a codimension-1 subvariety, which usually involves an increment in the degree of the ambient variety. By contrast, our strategy does not use any form of partitioning.

The main innovation of our work is a new method of relating degrees and orders of vanishing for multivariate polynomials. Earlier approaches, e.g., [27, 32, 33, 35, 38], consider multiple polynomials, and are related to understanding Bézout's theorem and possible inverses (see Tao's blog post [30] on inverse Bézout). Our approach instead only considers a single polynomial via parameter counting but we have to be extremely delicate in choosing vanishing conditions. We motivate and explain these ideas in Section 2. The polynomial method is already a powerful technique in discrete geometry, analysis, number theory, and theoretical computer science, and we hope that our method for handling higher dimensional objects will find additional applications.

The most general version of our result is Theorem 1.9 below, and it implies all the other statements. Next we gradually introduce the various generalizations and explain the history. The reader who is only interested in the proof of Theorem 1.2 can safely skip the rest of this section and proceed to Section 2 and Section 3 for the key ideas and the proof of Theorem 1.2.

1.1. **Joints of flats.** We extend Theorem 1.2 to flats of arbitrary dimensions. Given a collection of k-flats (i.e., k-dimensional flats) in $\mathbb{F}^{mk}$, a *joint* is defined to be a point contained in m of these k-flats and not all contained in a single hyperplane.

Theorem 1.3. *A set of N k-flats in $\mathbb{F}^{mk}$ has at most $C_{m,k}N^{m/(m-1)}$ joints, for some constant $C_{m,k}$.*

1.2. **Multijoints.** In the joints problem, instead of a single set of lines in $\mathbb{F}^d$, we can consider d sets of lines $\mathcal{L}_1, \dots, \mathcal{L}_d$ in $\mathbb{F}^d$ and consider joints formed by taking one line from each $\mathcal{L}_i$ (each point is counted as a joint at most once, for now). This variation, known as "multijoints", can be viewed as a discrete analogue of the endpoint multilinear Kakeya problem. The following bound on multijoints was conjectured by Carbery, proved in $\mathbb{F}^3$ and $\mathbb{R}^d$ by Iliopoulou [24] and in general $\mathbb{F}^d$ by Zhang [39]. Note that the the multijoints theorem is equivalent to the joints theorem if $|\mathcal{L}_i|$ are all within a constant factor of each other.

Theorem 1.4 (Multijoints of lines)**.** *Given d sets of lines $\mathcal{L}_1, \dots, \mathcal{L}_d$ in $\mathbb{F}^d$, the number of joints formed by taking one line from each $\mathcal{L}_i$ is at most $C_d(|\mathcal{L}_1| \cdots |\mathcal{L}_d|)^{1/(d-1)}$ for some constant C_d.*

We extend the multijoints theorem from lines to flats. Here a point is a *joint* formed by several flats if these flats contain this point and have spanning and independent directions.

Theorem 1.5 (Multijoints of flats)**.** *Given $\mathcal{F}_1, \dots, \mathcal{F}_r$, where $\mathcal{F}_i$ is a set of k_i-flats in $\mathbb{F}^d$, with $d = k_1 + \cdots + k_r$, the number of joints formed by taking one flat from each $\mathcal{F}_i$ is at most $C_{k_1,\dots,k_r}(|\mathcal{F}_1| \cdots |\mathcal{F}_r|)^{1/(r-1)}$ for some constant $C_{k_1,\dots,k_r}$.*

1.3. **Varieties.** We extend the joints theorem from flats to varieties. Generalizing earlier notions, a point p is a *joint* formed by several varieties $V_1, \dots, V_r$ if p is a regular point for each V_i and their tangent spaces at p have independent and spanning directions. (Recall that a point p is a *regular point* of a variety V if the Zariski tangent space T_pV has the same dimension as V.)

The proof of the joints theorem can be easily adapted from lines to algebraic curves (e.g., see [25, 26]). Here we extend the joints theorem to higher dimensional varieties. Given a set $\mathcal{V}$ of varieties, let $\deg \mathcal{V}$ denote the sum of the degrees of the elements of $\mathcal{V}$.

Theorem 1.6 (Joints of varieties)**.** *A set $\mathcal{V}$ of k-dimensional varieties in $\mathbb{F}^{mk}$ has at most $C_{m,k}(\deg \mathcal{V})^{m/(m-1)}$ joints for some constant $C_{m,k}$.*

Remark. In this paper, all varieties are assumed to be irreducible. We do not lose any generality for the joints problem with this assumption as one can always replace any algebraic set by its irreducible components.

Like earlier, we prove the result more generally for multiple sets of varieties.

Theorem 1.7 (Multijoints of varieties)**.** *Given $\mathcal{V}_1, \dots, \mathcal{V}_r$, where each $\mathcal{V}_i$ is a set of k_i-dimensional varieties in $\mathbb{F}^d$, where $d = k_1 + \cdots + k_r$, the number of joints formed by taking one variety from each V_i is at most $C_{k_1,\dots,k_r}(\deg \mathcal{V}_1 \cdots \deg \mathcal{V}_r)^{1/(r-1)}$ for some constant $C_{k_1,\dots,k_r}$.*

Previously, Iliopoulou [24] proved the multijoints theorem for algebraic curves of bounded degree in $\mathbb{R}^d$ (here by bounded degree we mean that the leading constant C depends on the maximum degree of the curves), but it was unknown how to to generalize from $\mathbb{R}^d$ to $\mathbb{F}^d$, despite knowledge of the joints theorem for a single set of curves. This is because Zhang's proof [39] of the multijoints theorem for lines (Theorem 1.4) does not easily adapt to curves.

In the setting of real varieties, Yang [35] proved an upper bound of the form $C_\epsilon(|\mathcal{V}_1| \cdots |\mathcal{V}_r|)^{1/(r-1)+\epsilon}$ for all $\epsilon > 0$ where C_ϵ also depends on the maximum degree of the varieties.

1.4. **Joints with multiplicities.** In the above formulations of joints and multijoints theorems, each point is counted as a joint at most once. Motivated by Kakeya problems, Carbery suggested a generalization where joints contained in many lines are counted with multiplicity. The following theorem about joints of lines with multiplicities was conjectured by Carbery, proved in $\mathbb{R}^3$ by Iliopoulou [21], and settled in general by Zhang [39].

Theorem 1.8 (Joints of lines with multiplicities)**.** *Let $\mathcal{L}_1, \dots, \mathcal{L}_d$ be multisets of lines in $\mathbb{F}^d$. Let $M(p)$ denote the number of tuples of lines $(\ell_1, \dots, \ell_d) \in \mathcal{L}_1 \times \cdots \times \mathcal{L}_d$ that form a joint at p. Summing over all such joints p, we have*

$$\sum_p M(p)^{1/(d-1)} \le C_d(|\mathcal{L}_1| \cdots |\mathcal{L}_d|)^{1/(d-1)},$$

where C_d is some constant.

Theorem 1.8 strengthens Theorem 1.4 (multijoints of lines). The exponent in $M(p)^{1/(d-1)}$ on the left-hand side is optimal as can be easily seen by duplicating every element in each set of lines m times for some large m.

Yang [35] studied a generalization of Theorem 1.8 to joints of varieties with multiplicities, but as earlier, his upper bound only holds in $\mathbb{R}^d$, carries an $+o(1)$ error term in the exponent, and the leading constant depends on the maximum degree of the varieties.

Our main result, below, generalizes the above to joints of varieties counted with multiplicities. It generalizes all previously stated results.

Theorem 1.9 (Joints of varieties with multiplicities)**.** *For each $i = 1, \dots, r$, let $\mathcal{V}_i$ be a multiset of k_i-dimensional varieties in $\mathbb{F}^d$, where $d = k_1 + \cdots + k_r$. Let $M(p)$ denote the number of tuples of varieties $(V_1, \dots, V_r) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_r$ that form a joint at p. Summing over all such joints p, we have*

$$\sum_p M(p)^{1/(r-1)} \le C_{k_1,\dots,k_r}(\deg \mathcal{V}_1 \cdots \deg \mathcal{V}_r)^{1/(r-1)},$$

where $C_{k_1,\dots,k_r}$ is some constant.

Our proof of Theorem 1.9 even in the case of lines is different from that of Zhang [39]. By our method, there is no significant difference between the proofs of Theorem 1.7 (without multiplicities) and Theorem 1.9 (with multiplicities).

1.5. **Constants.** We restate Theorems 1.7 and 1.9 in the following equivalent form with explicit constansts. This superficially more general formulation (formulated in [37] for flats) exposes a difficulty hierarchy of the problem. It also allows us to discuss the leading constants. While the constants below are optimal for $(r, k_1, m_1) = (1, 1, d)$, they are likely not tight in all other cases.

Theorem 1.10 (Main theorem)**.** *Let $k_1, \dots, k_r, m_1, \dots, m_r$ be positive integers. For each $i = 1, \dots, r$, let $\mathcal{V}_i$ be a finite multiset of k_i-dimensional varieties in $\mathbb{F}^d$, where $d = m_1k_1 + \cdots + m_rk_r$. We only consider joints p formed by choosing m_i unordered elements from $\mathcal{V}_i$ for each $i = 1, \dots, r$, and we write $M(p)$ for the number of such choices.*

(a) (without multiplicities) The number of joints is at most

$$C_{k_1,\dots,k_r;m_1,\dots,m_r}\left((\deg \mathcal{V}_1)^{m_1} \cdots (\deg \mathcal{V}_r)^{m_r}\right)^{1/(m_1+\cdots+m_r-1)},$$

where

$$C_{k_1,\dots,k_r;m_1,\dots,m_r} = \left(\frac{d!}{\prod_{i=1}^r k_i!^{m_i} m_i^{m_i}}\right)^{1/(m_1+\cdots+m_r-1)}.$$

(b) (with multiplicities) Summing over all joints p, one has

$$\sum_p M(p)^{1/(m_1+\cdots+m_r-1)} \le C'_{k_1,\dots,k_r;m_1,\dots,m_r}\left((\deg \mathcal{V}_1)^{m_1} \cdots (\deg \mathcal{V}_r)^{m_r}\right)^{1/(m_1+\cdots+m_r-1)},$$

where

$$C'_{k_1,\dots,k_r;m_1,\dots,m_r} = \left(\frac{d!}{\prod_{i=1}^r k_i!^{m_i} m_i!}\right)^{1/(m_1+\cdots+m_r-1)}.$$

Let us explain how various specializations of Theorem 1.10 correspond to earlier results.

(1) (Joints of lines) Theorem 1.1 corresponds to Theorem 1.10(a) for $r = 1$, $k_1 = 1$, $m_1 = d$, and degree 1 varieties. In this case, the optimal constant $C_{1;d} = (d-1)!^{1/(d-1)}/d$ was determined previously in [37] and matches the constant above.

(2) (Joints of flats) Theorem 1.3 corresponds to Theorem 1.10(a) for $r = 1$, $k_1 = k$, $m_1 = m$, $d = km$, and degree 1 varieties.

(3) (Multijoints of lines) Theorem 1.4 corresponds to Theorem 1.10(a) with $r = d$, $(k_i, m_i) = (1, 1)$ for all i, and degree 1 varieties. This case was previously known [39]. The constant in this case was improved to $C_{1,\dots,1;1,\dots,1} = d!^{1/(d-1)}$ in [37] (matching above) though it is likely not optimal. Without consideration of constants, Theorem 1.4 also easily implies the setting allowing $m_i \geq 1$ by duplicating the sets of lines.

(4) (Multijoints of k-flats) Theorem 1.5 relaxes the $k_i = 1$ assumption above to arbitrary $k_i \geq 1$. Previously the only other known case is $(r; k_1, k_2; m_1, m_2) = (2; k, 1; 1, d-k)$, i.e., a set of k-flats and a set of lines, where each joint is formed by one k-flat and $d-k$ lines, as proved independently by [5] and [37] (and stated without proof in [35]). Even the "next" case of $(r; k_1, k_2; m_1, m_2) = (2; 2, 1; 2, 1)$ was previously unknown, corresponding to having a joint being formed by two flats and one line. Likewise, the case $(r; k_1, k_2, k_3; m_1, m_2, m_3) = (3; 2, 1, 1; 1, 1, 1)$ allowing one set of flats and two different sets of lines was also previously unsolved.

(5) (Varieties) Theorems 1.6 and 1.7 relax the degree 1 assumption, generalizing from flats to varieties. The only previously known case was for a single set of curves [25, 26], namely $r = 1$ and $k_1 = 1$, as well as multiple sets of bounded degree curves in $\mathbb{R}^n$ [24]. Theorem 1.7 is equivalent to Theorem 1.10(a) (other than constants).

(6) (Multiplicities) Finally, adding in considerations of joint multiplicities, Theorem 1.8 is equivalent to Theorem 1.10(b) for lines, while Theorem 1.9 is equivalent to Theorem 1.10(b) in general (other than constants). For a single set of lines, i.e., $(r, k_1, m_1) = (1, 1, d)$, our result gives $C_{1;d} = 1$.

While we know the optimal constant for joints of lines, our proof does not seem to give the optimal constant for flats or varieties. For $r = 1$ we conjecture that the optimal constant in Theorem 1.10 is $C_{k,m} = (m!/m^m)^{1/(m-1)} N^{m/(m-1)}$ for all k and m, agreeing with joints of lines ($k = 1$). The first open case $(k, m) = (2, 3)$ is stated below.

Conjecture 1.11. *A set of N planes in $\mathbb{F}^6$ has at most $(\sqrt{2}/3 + o(1))N^{3/2}$ joints.*

1.6. **Outline.** We begin by motivating and describing, in Section 2, the key new ideas in our method. We then give, in Section 3, the proof in the special case of joints of planes in $\mathbb{R}^6$, which is representative of the general result. To obtain the result in full generality, we use higher order directional derivatives with respect to local coordinates along a variety, as well as Hasse derivatives to deal with arbitrary fields, and they are both discussed in Section 4. The complete proof of the main theorem then appears in Section 5.

2. Key ideas

2.1. **Joints of lines.** We begin by recalling the proof of Theorem 1.1 on joints of lines in $\mathbb{R}^3$ following [25, 26] (also see Guth's book [15, Section 2.5] for a nice exposition). The proof exposes two tools that are essential in nearly all applications of the polynomial method: *parameter counting* and *vanishing lemma.*

Let $\mathbb{R}[x_1,\dots,x_d]_{\le n}$ denote the space of polynomials with degree at most n. Using that its dimension is $\binom{n+d}{d}$, we have the following simple yet extremely useful linear algebraic consequence.

Lemma 2.1 (Parameter counting)**.** *Given a set of fewer than $\binom{n+3}{3}$ points in $\mathbb{R}^3$, there exists a nonzero polynomial of degree at most n that vanishes on all these points.*

Given a set of N lines forming J joints in $\mathbb{R}^n$, let g be a nonzero polynomial of minimum degree that vanishes at all J joints. By the parameter counting lemma, we have $\deg g \le CJ^{1/3}$ for some constant $C > 0$.

The following elementary fact is key to the polynomial method.

Lemma 2.2 (Vanishing lemma)**.** *If a degree n polynomial vanishes at more than n points on a line, then it vanishes on the whole line.*

We claim that some line contains at most $CJ^{1/3}$ joints. Suppose, for contradiction, that every line contains more than $CJ^{1/3}$ joints. Since $\deg g \le CJ^{1/3}$, the vanishing lemma implies that g vanishes on each of the N lines. Since each joint is contained in three lines in spanning directions, the gradient ∇g vanishes at every joint. Thus $\partial g/\partial x$, $\partial g/\partial y$, $\partial g/\partial z$ all vanish at every joint. At least one of these partial derivatives is a nonzero polynomial of degree smaller than that of g, thereby contradicting the minimal degree assumption on g.

Thus some line contains at most $CJ^{1/3}$ joints. We can then remove this line and all its joints, and repeat the argument to find another line with at most $CJ^{1/3}$ joints. After we have removed all the lines, we have removed at most $CJ^{1/3}N$ joints, so $J \le CJ^{1/3}N$, and hence $J = O(N^{3/2})$. This completes the proof in the case of $\mathbb{R}^3$. This proof also extends to $\mathbb{F}^d$.

2.2. **Vanishing on planes.** How can we try to adapt the above proof to show that N planes in $\mathbb{R}^6$ form $O(N^{3/2})$ joints? The main obstacle is to generalize the vanishing lemma from lines to planes. The above proof would extend verbatim to joints of planes if the answer to the following question were yes.

Attempt I. Given distinct points $p_1,\dots,p_{\binom{n+2}{2}}$ in the plane, if $g \in \mathbb{R}[x,y]_{\le n}$ satisfies the vanishing conditions $g(p_1) = 0,\dots,g(p_{\binom{n+2}{2}}) = 0$, does this imply that g is identically zero?

Of course, the answer to this question is no, since the vanishing locus of the polynomial on a plane could be a curve. Clearly it is impossible to force a two-variable polynomial to vanish by forcing it to vanish at any finite number of points. Instead of asking for polynomials to vanish at the joints, we can ask them to vanish to high multiplicity at the joints. This idea, known as the "method of multiplicities" [10], has been fruitful in the study of the joints problem [39, 37], and it was also used to improve bounds on the finite field Kakeya problem [8, 3].

Attempt II. Given a point p_1 in the plane, if $g \in \mathbb{R}[x,y]_{\le n}$ vanishes to order more than n at p_1; equivalently, if g satisfies the vanishing conditions $\frac{\partial^{i+j}g}{\partial x^i \partial y^j}(p_1) = 0$ for all $0 \le i+j \le n$, does this imply that g is identically zero?

The answer to this one is yes, and it shows how using derivatives creates a correct vanishing lemma. However, this vanishing lemma is completely useless for our application since we want to use this vanishing lemma somehow to bound the number of joints lying on a plane and this method ignores all of the joints but one on each plane. Perhaps we can create a correct and useful vanishing lemma by combining the ideas of Attempts I and II.

Attempt III. Given distinct points $p_1,\dots,p_m$ in the plane with $m \sim n^2/s^2$, if $g \in \mathbb{R}[x,y]_{\le n}$ vanishes to order at least s at each point, does this imply that g is identically zero?

Unfortunately the answer is no again. Indeed $g(x,y) = y^s$ vanishes to order s on the entire x-axis.

We have $\dim \mathbb{R}[x,y]_{\leq n} = \binom{n+2}{2}$, less than the number of linear constraints on the coefficients of g imposed by asking g to vanish to order at least s on $\Theta(n^2/s^2)$ given points (each such point gives $\binom{s+2}{2}$ constraints). This counterexample must imply that some of these linear constraints are linearly dependent. Our proof strategy is to build a vanishing lemma using a linearly independent set of such constraints on the coefficients of g.

Remark. Another very natural strategy for extending the proof of joints of lines to planes is to consider, instead of a single polynomial that vanishes on all the joints, now a pair of polynomials that vanish on all the joints. For this approach to be useful, one would like the pair of polynomials, when restricted to each plane, to either be coprime or one of them to vanish. This seems like a difficult condition to satisfy and we suspect that it is not possible, at least if one wants the degrees of the polynomials to be small.

This problem appears to be related to the *inverse Bézout problem.* Given a set of N points in $\mathbb{R}^2$, can one always find a pair of coprime polynomials P, Q both vanishing on all N points and $(\deg P)(\deg Q) = O(N)$? The answer is no, by putting half of the N points on a $\sqrt{N/2} \times \sqrt{N/2}$ grid and the other half on a line (this grid-and-line example shows up again in our discussion below). A partial converse to Bézout's theorem is known in 2-dimensions but open in higher dimensions (see Tao [30]).

2.3. **Key idea I: collecting linearly independent vanishing conditions.** We define a *vanishing condition* to be a single homogeneous linear constraint on the coefficients of a polynomial $g \in \mathbb{R}[x,y]_{\leq n}$ that arises from requiring some particular higher order directional derivative to vanish at some point. For example, for a two variable polynomial g, some examples of vanishing conditions are (a) $g(2,4) = 0$, (b) $\frac{\partial g}{\partial x}(2,1) = 0$, and (c) $\left(\frac{\partial^2 g}{\partial x^2} - \frac{\partial^2 g}{\partial x \partial y}\right)(-1,2) = 0$. For a positive integer r, an *r-th order vanishing condition* on g at p is a vanishing condition of the form $Dg(p) = 0$ where D is an $(r-1)$-th order derivative operator, i.e., a linear combination of $\partial^{r-1}/\partial^{r_1} x_1 \cdots \partial^{r_d} x_d$ for some $r_1 + \cdots + r_d = r - 1$. (We will not need mixed order vanishing conditions for joints of flats, but they will be needed for joints of varieties.)

For now, let us focus on a single plane and study vanishing conditions on $g \in \mathbb{R}[x,y]_{\leq n}$. Vanishing conditions can be viewed as linear functionals on the vector space $\mathbb{R}[x,y]_{\leq n}$, though it will be helpful later to also keep track of the (derivative operator, point) pair (D, p) that generates the vanishing condition $Dg(p) = 0$.

We now devise a procedure for selecting a basis of linear functionals on $\mathbb{R}[x,y]_{\leq n}$.

As a first attempt, we fix an arbitrary order on $\mathcal{P}$, say $p_1, \ldots, p_r$ and cycle through the points (the vertical bars are a visual aid separating the epochs)

$$p_1 \ p_2 \ \cdots p_r \mid p_1 \ p_2 \ \cdots p_r \mid p_1 \ p_2 \ \cdots p_r \mid \cdots .$$

We cycle through the points in the above sequence and maintain a linearly independent set of vanishing conditions on $\mathbb{R}[x,y]_{\leq n}$, starting from an empty set of vanishing conditions. The r-th time ($r = 1, 2, \ldots$) that we see a point p, we add to our existing collection a maximal subset of r-th order vanishing conditions so that our collection of vanishing conditions always remains linearly independent as a set of linear functionals on $\mathbb{R}[x,y]_{\leq n}$. Eventually, the process terminates once we have collected a basis of $\binom{n+2}{2}$ linear functionals on $\mathbb{R}[x,y]_{\leq n}$.

Although there is some choice in the above process in deciding which vanishing conditions to add to our collection at each step, the *number* of vanishing conditions added at each step does not depend on this choice. We would like to understand and control the number of vanishing conditions attached to each point as we run through the process. However, this does not seem easy. We do not know how to compute these numbers (for large n) even for an explicitly given set of points.

More importantly, the process does not always evenly assign the vanishing conditions across all the points. For example, suppose we have $|\mathcal{P}| = 2t^2$, with half of the points in $\mathcal{P}$ forming an $t \times t$ grid (a high-degree part), and the other t^2 points all lying on a single generic line (a low-degree part). As we run through the above process, we encounter significantly more linear dependencies among vanishing conditions at points on the line than on the grid. For large n, at the end of the process, each point on the grid receives on the order of t times as many vanishing conditions as each point on line. This is an undesirable situation, since the process leads to an unequal distribution of vanishing conditions, effectively "ignoring" the points on the low-degree algebraic structure.

2.4. **Key idea II: handicaps and priority order.** To address the uneven distribution of vanishing conditions across points, we give the "disadvantaged" points a head start and cycle just among themselves many times before we cycle through the entire set of points. For example, in the earlier grid-and-line example, if $p_1, \dots, p_{r/2}$ are points on the line and $p_{r/2+1}, \dots, p_r$ are points on the grid, then we give points on the line a head start, e.g.,

$$p_1\ p_2 \cdots p_{r/2} \mid \cdots \mid p_1\ p_2 \cdots p_{r/2} \mid p_1\ p_2 \cdots p_r \mid p_1\ p_2 \cdots p_r \mid \cdots .$$

More generally, we give each point p a *handicap* $\alpha_p \in \mathbb{Z}$ corresponding to the number of rounds of head start.

For example, suppose there are five points labeled a, b, c, d, e that we would cycle through in this order. Now we assign handicaps $0, 1, 3, 0, -1$ to a, b, c, d, e respectively. Then, for instance, c starts in round -3 and b starts in round -1. So we process the points in the following *priority order*:

$$c \mid c \mid b\ c \mid a\ b\ c\ d\ \mid a\ b\ c\ d\ e \mid a\ b\ c\ d\ e \mid \cdots .$$

We now run the same vanishing condition collection process as earlier with this sequence of points. The r-th time ($r = 1, 2, \dots$) that we see a point p, we append to our existing collection a maximal non-redundant set of r-th order vanishing conditions at p.

We would like to assign handicaps in a way so that all joints are treated equitably in the distribution of vanishing conditions (what this means precisely will be explained later). However, it appears to be a very difficult problem to determine how exactly the distribution of vanishing conditions depends on the handicaps. Intuitively, as in the grid-and-line example, we want to assign more handicap to points that are part of low-degree algebraic substructures, but it is far from obvious how to make this notion precise and useful.

2.5. **Key idea III: existence of a good handicap via compactness/smoothing.** Instead of explicitly assigning handicaps, we shall indirectly prove the existence of a good choice of handicaps via a compactness/smoothing argument. (Strictly speaking, we do not actually invoke compactness here since all our domains are finite, but we believe that compactness offers a helpful perspective as the argument here is a significant generalization of the earlier compactness argument giving tight bounds for joints of lines [37].)

Fix a joints configuration. Let n be large and consider the function

$$\text{handicaps } \alpha \in \mathbb{Z}^{\mathcal{P}} \longrightarrow \text{partitions of } \binom{n+2}{2} \text{ among } \mathcal{P} \tag{2.1}$$

where the partition records the final number of vanishing conditions assigned to each point. While it appears to be difficult to compute this function explicitly, we can show that it has the following three properties.

Bounded domain. If one point has a much bigger handicap than another point, then the latter point gets assigned no vanishing conditions since the process would have finished before the first appearance of the latter point. Such a situation will never be desirable, so we only need to consider cases where the handicaps are all bounded (as a function of n).

Monotonicity. Suppose we increase the handicap by one at a subset of points while holding others fixed. Then the number of vanishing conditions assigned to this subset of points cannot decrease, and the number of vanishing conditions assigned to the other points cannot increase. Indeed, the points with the increased handicap now appear earlier in the priority order, and thus cannot receive fewer vanishing conditions than before the change.

Lipschitz continuity. A small change in the handicap assignments can only induce a small change in the number of vanishing conditions at each point. This property is intuitively reasonable, but it requires a proof.

With these three properties, we can iteratively increase the handicaps at points that end up with too few constraints, so that we eventually balance out the distribution of constraints across all joints.

The eventual implicit assignment of handicaps across joints appears to somehow identify the "algebraicity" of each point in the configuration by assigning higher handicaps to points lying in lower-degree algebraic substructures. However, we do not know how to make this algebraicity intuition precise.

Remark. This idea of implicitly assigning handicaps came up in a simpler form previously in the work of Yu and Zhao [37] in determining the tight constant for the joints theorem of lines. There one does not have to consider any priority order or iterative process of adding constraints as we do here, though one does end up proving, via compactness, the existence of a handicap (though not called by that name) along with other parameters for controlling the order of vanishing at each joint.

2.6. **Putting everything together: a new vanishing lemma.** Suppose we have a set $\mathcal{F}$ of planes in $\mathbb{R}^6$ forming joints $\mathcal{J}$. For a choice of handicaps $\vec{\alpha} \in \mathbb{Z}^{\mathcal{J}}$, and a large integer n, we can run the above vanishing condition collection procedure separately on each plane (using handicaps $\vec{\alpha}$ restricted to points on the plane). On each plane $F \in \mathcal{F}$, and at each joint p on the plane F, the procedure attaches a set $\mathcal{D}_{p,F} = \mathcal{D}_{p,F}(\vec{\alpha}, n)$ of derivative operators. Combining these vanishing conditions over all joints on F then gives a basis of linear functionals on the space of polynomials g on F of degree at most n, where each basis element is a vanishing condition of the form $Dg(p) = 0$ with $p \in \mathcal{F}$ and $D \in \mathcal{D}_{p,F}$ being a linear combination of higher order directional derivatives along F. With this data, we can now state our new vanishing lemma for joints of planes.

Vanishing lemma for joints of planes (Lemma 3.9). *With the above setup, if $g \in \mathbb{R}[x_1, \ldots, x_6]_{\le n}$ satisfies $D_1 D_2 D_3 g(p) = 0$ whenever $D_i \in \mathcal{D}_{p,F_i}$ are three derivative operators attached to three planes F_1, F_2, F_3 forming a joint $p \in \mathcal{J}$, then $g = 0$.*

Note that we are choosing a minimal set of derivative operators on each plane (as we chose a basis of linear functionals). The vanishing lemma would be trivial if each $\mathcal{D}_{p,F_i}$ were the full set of directional derivative operators at p along F.

Also, our proof of the vanishing lemma only works if we build the vanishing conditions following the priority order—we would not be able to say much if the joints were processed in some other arbitrary manner.

By **parameter counting**, this new vanishing lemma implies the following inequality. Summing over joints p formed by a triple of planes F_1, F_2, F_3, we have

$$\sum_{(p,F_1,F_2,F_3)} |\mathcal{D}_{p,F_1}|\,|\mathcal{D}_{p,F_2}|\,|\mathcal{D}_{p,F_3}| \ge \dim \mathbb{R}[x_1, \ldots, x_6]_{\le n} = \binom{n+6}{6}.$$

The left-hand side is the number of linear constraints on g of the form $D_1 D_2 D_3 g(p) = 0$ in the vanishing lemma. Indeed, if this inequality were not satisfied, by parameter counting there would be a non-zero polynomial g of degree at most d satisfying these vanishing conditions. However, the vanishing lemma implies that such a g is identically zero, a contradiction.

Recall that all these quantities $|\mathcal{D}_{p,F}|$ depend on n as well as the handicap $\vec{\alpha}$. We can now apply a compactness/smoothing argument to choose a handicap $\vec{\alpha}$ that minimizes

$$\max_p |\mathcal{D}_{p,F_1}|\,|\mathcal{D}_{p,F_2}|\,|\mathcal{D}_{p,F_3}| - \min_p |\mathcal{D}_{p,F_1}|\,|\mathcal{D}_{p,F_2}|\,|\mathcal{D}_{p,F_3}|\,.$$

Using the three properties (bounded domain, monotonicity, Lipschitz continuity) of (2.1), we can deduce that the above difference must be negligible, i.e., $o(n^6)$, since otherwise we can significantly reduce the above difference by increasing the handicap by 1 at a subset of points p with small $|\mathcal{D}_{p,F_1}|\,|\mathcal{D}_{p,F_2}|\,|\mathcal{D}_{p,F_3}|$.

It follows that we can choose handicaps so that the product $|\mathcal{D}_{p,F_1}|\,|\mathcal{D}_{p,F_2}|\,|\mathcal{D}_{p,F_3}|$ is roughly constant across all (p, F_1, F_2, F_3). We also know that for each plane F, $\sum_{p\in F} |\mathcal{D}_{p,F}| = \dim \mathbb{R}[x, y]_{\le n} = \binom{n+2}{2}$ since we have a basis of linear functionals on the space of polynomials on F with degree at most n. The conclusion $|\mathcal{J}| = O(N^{3/2})$ then follows from a short calculation using the AM-GM inequality (see the end of Section 3).

In Section 3, we flesh out these ideas to give a complete proof of joints of planes in $\mathbb{R}^6$. In Section 4 we discuss two further modifications to the above proof technique. To deal with varieties, we modify our notion of higher order directional derivatives. Geometrically we are taking derivatives with respect to local coordinates on the varieties. To deal with general fields other than the reals, we use Hasse derivatives.

3. Joints of planes in $\mathbb{R}^6$

The purpose of this section is to prove that N planes in $\mathbb{R}^6$ have $O(N^{3/2})$ joints. This special case contains many of the key ideas that we introduce in this paper towards the full theorem.

Let $(\mathcal{J}, \mathcal{F})$ be a *joints configuration* of planes in $\mathbb{R}^6$, where $\mathcal{F}$ is a finite set of planes and $\mathcal{J}$ is the set of joints formed by any three planes in $\mathcal{F}$. We abuse notation slightly to handle the case when more than three planes pass through $p \in \mathcal{J}$: in this case we arbitrarily choose three planes forming a joint at p, and only write "$p \in F$" (and say that "F contains p", etc.) if F is among the triple of planes chosen at p.

3.1. **Priority order and handicaps.**

First, assign an arbitrary but fixed order (referred to as the *preassigned order*) to the joints $\mathcal{J}$.

A *handicap* $\vec{\alpha} = (\alpha_p)_{p\in\mathcal{J}} \in \mathbb{Z}^{\mathcal{J}}$ assigns an integer to each joint. Given a handicap, the associated *priority order* is a linear order on $\mathcal{J} \times \mathbb{Z}_{\ge 0}$ defined by setting $(p, r) \prec (p', r')$

- if $r - \alpha_p < r' - \alpha_{p'}$, or
- if $r - \alpha_p = r' - \alpha_{p'}$ and p comes before p' in the preassigned order on $\mathcal{J}$.

The priority ordering corresponds to the description in the previous section. Note that in particular $(p, 0) \prec (p, 1) \prec (p, 2) \prec \cdots$. We write $\prec$ for the strict ordering, and $\preceq$ to allow equality.

3.2. **Derivatives and evaluations.**

Let $\mathbb{R}[x_1, \dots, x_k]_{\le n}$ denote the space of polynomials of degree at most n in k variables.

Given a plane F and a joint $p \in F$, let $\mathbb{D}^r_{p,F}$ denote the space of all r-th order derivative operators in directions along F, i.e., every element $D \in \mathbb{D}^r_{p,F}$ gives a linear map $g \mapsto Dg$ sending $\mathbb{R}[x_1 \dots, x_6] \to \mathbb{R}[x_1, \dots, x_6]$ and D is a linear combination of compositions of r directional derivative operators along F. For example, if F is the plane spanned by the first two coordinate directions, then $\mathbb{D}^r_{p,F}$ is the space spanned by the operators $\partial^{i+j}/\partial x_1^i \partial x_2^j$ ranging over all $i + j = r$. (The space $\mathbb{D}^r_{p,F}$ here does not actually depend on p, but we include p in the notation with a view towards generalization from flats to varieties.)

Let $\mathbb{B}^r_{p,F}(n)$ denote the subspace of all linear functionals on $\mathbb{R}[x_1, \dots, x_6]_{\le n}$ of the form $g \mapsto Dg(p)$ for some $D \in \mathbb{D}^r_{p,F}$ (i.e., an r-th order derivative along F evaluated at p). Then, for a fixed $p \in \mathcal{J} \cap F$, a polynomial $g \in \mathbb{R}[x_1, \dots, x_6]_{\le n}$ lies in the common kernel of $\mathbb{B}^0_{p,F}(n) + \mathbb{B}^1_{p,F}(n) + \cdots + \mathbb{B}^{r-1}_{p,F}(n)$ if and only if the restriction of g to the plane F vanishes to order at least r at p. (By common kernel we mean the intersection of the kernels of all linear functionals in this space.)

To emphasize the difference between $\mathbb{B}$ and $\mathbb{D}$, the elements of $\mathbb{D}^r_{p,F}$ are derivative operators sending polynomials to polynomials, whereas the elements of $\mathbb{B}^r_{p,F}(n)$ are linear functionals sending polynomials of degree up to n to scalars. Perhaps a helpful mnemonic is that D stands for "differentiation" while B stands for "basis" (we will soon use a basis of the space of linear forms on polynomials up to degree n).

For a fixed $F \in \mathcal{F}$, let us describe a process where we go through pairs $(p, r) \in (\mathcal{J} \cap F) \times \mathbb{Z}_{\geq 0}$ according to the priority order, and at each step we choose a

$$\mathcal{B}^r_{p,F}(\vec{\alpha}, n) \subset \mathbb{B}^r_{p,F}(n).$$

We will drop the dependencies on $\vec{\alpha}$, n, and F when there is no confusion, i.e., we write $\mathcal{B}^r_p \subset \mathbb{B}^r_p$ for the above inclusion. In addition, all unions and direct sums in the following paragraph are taken over $(p', r') \in (\mathcal{J} \cap F) \times \mathbb{Z}_{\geq 0}$.

Suppose we are at the start of step (p, r). At this point, we have already chosen some $\mathcal{B}^{r'}_{p'} \subset \mathbb{B}^{r'}_{p'}$ for each $(p', r') \prec (p, r)$ so that the disjoint union $\bigcup_{(p',r')\prec(p,r)} \mathcal{B}^{r'}_{p'}$ is a basis for $\sum_{(p',r')\prec(p,r)} \mathbb{B}^{r'}_{p'}$. Now consider expanding this space to $\sum_{(p',r')\preceq(p,r)} \mathbb{B}^{r'}_{p'}$ by adding in all the r-th order derivative evaluations at p along F. We desire to expand the basis accordingly. As such, we choose a set $\mathcal{B}^r_p \subset \mathbb{B}^r_p$ so that the disjoint union $\bigcup_{(p',r')\preceq(p,r)} \mathcal{B}^{r'}_{p'}$ becomes a basis of $\sum_{(p',r')\preceq(p,r)} \mathbb{B}^{r'}_{p'}$. Note that while we have some choice about which elements of $\mathbb{B}^r_p$ to include as new basis elements, the size of $\mathcal{B}^r_p$ does not depend on any choice, and is only a function n and the priority order. We will provide a more direct formula for $\left|\mathcal{B}^r_p\right|$ shortly.

Since each element of $\mathbb{B}^r_{p,F}(n)$ can be written as $g \mapsto Dg(p)$ for some $D \in \mathbb{D}^r_{p,F}$, we can choose

$$\mathcal{D}^r_{p,F}(\vec{\alpha}, n) \subset \mathbb{D}^r_{p,F}$$

with the same size as $\mathcal{B}^r_{p,F}(\vec{\alpha}, n)$ so that

$$\mathcal{B}^r_{p,F}(\vec{\alpha}, n) = \{g \mapsto Dg(p) : D \in \mathcal{D}^r_{p,F}(\vec{\alpha}, n)\}.$$

We write

$$\mathcal{B}_{p,F}(\vec{\alpha}, n) := \bigcup_{r\geq 0} \mathcal{B}^r_{p,F}(\vec{\alpha}, n) \quad \text{and} \quad \mathcal{D}_{p,F}(\vec{\alpha}, n) := \bigcup_{r\geq 0} \mathcal{D}^r_{p,F}(\vec{\alpha}, n).$$

As we range over all joints p on F, the sets $\mathcal{B}_{p,F}(\vec{\alpha}, n)$ combine to form a basis of the space of linear forms on polynomials of degree at most n on F. Thus

$$\sum_{p\in\mathcal{J}\cap F} |\mathcal{B}_{p,F}(\vec{\alpha}, n)| = \dim \mathbb{R}[x, y]_{\leq n} = \binom{n+2}{2}. \tag{3.1}$$

We may omit the parenthetical $\vec{\alpha}$ and n in our notation when these parameters do not change and the context is clear. Some of the arguments below will involve comparing different values of $\vec{\alpha}$ and n, in which case we will state the dependencies explicitly. We may also omit F when we are not considering other planes.

3.3. **Polynomials with given vanishing orders.** In this and the next subsection, we focus our attention on a single fixed plane $F \cong \mathbb{R}^2$. Fix a finite set of points $\mathcal{P} \subset F$ (which we will later take to be the joints on F). Given a vector $\vec{v} = (v_p)_{p\in\mathcal{P}} \in \mathbb{Z}^{\mathcal{P}}_{\geq 0}$, let

$$\mathbb{T}(\vec{v}, n) = \{g \in \mathbb{R}[x, y]_{\leq n} : g \text{ vanishes to order} \geq v_p \text{ at each } p \in \mathcal{P}\}$$

(i.e., the partial derivatives satisfy $\frac{\partial^{i+j} g}{\partial x^i \partial y^j}(p) = 0$ for all $i + j < v_p$). We would like to understand how the dimension of $\mathbb{T}(\vec{v}, n)$ changes with $\vec{v}$ and n. We are particularly interested in the following quantity, which we will shortly relate below in (3.4) to $|\mathcal{B}^r_{p,F}(\vec{\alpha}, n)|$: for $p \in \mathcal{P}$, set

$$b_p(\vec{v}, n) := \operatorname{codim}_{\mathbb{T}(\vec{v},n)} \mathbb{T}(\vec{v} + \vec{e}_p, n) = \dim \mathbb{T}(\vec{v}, n) - \dim \mathbb{T}(\vec{v} + \vec{e}_p, n).$$

Here, given a pair of subspaces $W \leq U$, we write $\operatorname{codim}_U W$ for the relative codimension of W in U. Also $\vec{e}_p \in \mathbb{Z}^{\mathcal{P}}$ is the vector with 1 at p and 0 elsewhere. Note, for each $p \in \mathcal{P}$, the space $\mathbb{T}(\vec{v} + \vec{e}_p, n)$ is the nullspace of the map on $\mathbb{T}(\vec{v}, n)$ that sends every polynomial g to all its v_p-th order derivatives evaluated at p, and thus $b_p(\vec{v}, n)$ is the rank of this map.

The following basic fact will be useful:

$$\text{for subspaces } U, W \le V, \text{ we have } \operatorname{codim}_V W \ge \operatorname{codim}_U (W \cap U). \tag{3.2}$$

Lemma 3.1 (Bounded domain)**.** *If $\vec{v} \in \mathbb{Z}_{\ge 0}^{\mathcal{P}}$ has $v_p > n$ for some $p \in \mathcal{P}$, then $\dim \mathbb{T}(\vec{v}, n) = 0$.*

Proof. This is the statement that no nonzero polynomial of degree at most n can vanish to order more than n at some point. □

Lemma 3.2 (Monotonicity)**.** *Let $p \in \mathcal{P}$. Suppose $\vec{v}^{(1)}, \vec{v}^{(2)} \in \mathbb{Z}_{\ge 0}^{\mathcal{P}}$ satisfy $\vec{v}^{(1)} \ge \vec{v}^{(2)}$ coordinatewise and with equality at p. Then $b_p(\vec{v}^{(1)}, n) \le b_p(\vec{v}^{(2)}, n)$ for all n.*

Proof. Earlier we saw that for each $i = 1, 2$, $b_p(\vec{v}^{(i)}, n)$ is the rank of the map on $\mathbb{T}(\vec{v}^{(i)}, n)$ that sends each polynomial to all its $v_p^{(1)} = v_p^{(2)}$-th order derivatives evaluated at p. Since $\vec{v}^{(1)} \ge \vec{v}^{(2)}$ coordinatewise, $\mathbb{T}(\vec{v}^{(1)}, n)$ is a subspace of $\mathbb{T}(\vec{v}^{(2)}, n)$, which implies the inequality $b_p(\vec{v}^{(1)}, n) \le b_p(\vec{v}^{(2)}, n)$ on the rank of a map when restricted to a subspace. □

The next two lemmas together will lead to the Lipschitz continuity property of $b_p(\vec{v}, n)$ as a function of $\vec{v}$.

Lemma 3.3. *Let $p, q \in \mathcal{P}$ be distinct points. Then for every $\vec{v} \in \mathbb{Z}_{\ge 0}^{\mathcal{P}}$ and nonnegative integer n, one has $b_p(\vec{v} + \vec{e}_q, n) \ge b_p(\vec{v}, n-1)$.*

Proof. Let f be an arbitrary linear polynomial that vanishes at q but at no other point of $\mathcal{P}$ (such f clearly exists if the underlying field $\mathbb{F}$ is large enough; if not, we replace $\mathbb{F}$ by a field extension, which would not affect $b_p(\vec{v}, n)$ as it is a rank-type quantity). We have

$$\begin{aligned} b_p(\vec{v} + \vec{e}_q, n) &= \operatorname{codim}_{\mathbb{T}(\vec{v}+\vec{e}_q, n)} \mathbb{T}(\vec{v} + \vec{e}_p + \vec{e}_q, n) \\ &\ge \operatorname{codim}_{f \cdot \mathbb{T}(\vec{v}, n-1)} f \cdot \mathbb{T}(\vec{v} + \vec{e}_p, n-1) \\ &= \operatorname{codim}_{\mathbb{T}(\vec{v}, n-1)} \mathbb{T}(\vec{v} + \vec{e}_p, n-1) \\ &= b_p(\vec{v}, n-1). \end{aligned}$$

The inequality step follows from (3.2), observing that restricting $\mathbb{T}(\vec{v} + \vec{e}_q, n)$ and $\mathbb{T}(\vec{v} + \vec{e}_p + \vec{e}_q, n)$ to polynomials divisible by f yields $f \cdot \mathbb{T}(\vec{v}, n-1)$ and $f \cdot \mathbb{T}(\vec{v} + \vec{e}_p, n-1)$ respectively. □

Lemma 3.4. *Let $p \in \mathcal{P}$. Suppose $\vec{v}^{(0)}, \vec{v}^{(1)}, \dots \in \mathbb{Z}^{\mathcal{P}}$ are such that $\vec{v}^{(0)} \le \vec{v}^{(1)} \le \cdots$ coordinate-wise and strictly increasing at the coordinate indexed by p. Then*

$$\sum_{r \ge 0} b_p(\vec{v}^{(r)}, n) - \sum_{r \ge 0} b_p(\vec{v}^{(r)}, n-1) \le n + 1.$$

Proof. For each $r \ge 0$, we have

$$\begin{aligned} b_p(\vec{v}^{(r)}, n) - b_p(\vec{v}^{(r)}, n-1) &= \operatorname{codim}_{\mathbb{T}(\vec{v}^{(r)}, n)} \mathbb{T}(\vec{v}^{(r)} + \vec{e}_p, n) - \operatorname{codim}_{\mathbb{T}(\vec{v}^{(r)}, n-1)} \mathbb{T}(\vec{v}^{(r)} + \vec{e}_p, n-1) \\ &= \operatorname{codim}_{\mathbb{T}(\vec{v}^{(r)}, n)} \mathbb{T}(\vec{v}^{(r)}, n-1) - \operatorname{codim}_{\mathbb{T}(\vec{v}^{(r)}+\vec{e}_p, n)} \mathbb{T}(\vec{v}^{(r)} + \vec{e}_p, n-1). \end{aligned}$$

We have $\operatorname{codim}_{\mathbb{T}(\vec{v}^{(r)}+\vec{e}_p, n)} \mathbb{T}(\vec{v}^{(r)} + \vec{e}_p, n-1) \ge \operatorname{codim}_{\mathbb{T}(\vec{v}^{(r+1)}, n)} \mathbb{T}(\vec{v}^{(r+1)}, n-1)$ by (3.2) since $\vec{v}^{(r)} + \vec{e}_p \le \vec{v}^{(r+1)}$ coordinatewise. Summing over all $r \ge 1$, we obtain

$$\begin{aligned} \sum_{r \ge 0} b_p(\vec{v}^{(r)}, n) - \sum_{r \ge 0} b_p(\vec{v}^{(r)}, n-1) &\le \operatorname{codim}_{\mathbb{T}(\vec{v}^{(0)}, n)} \mathbb{T}(\vec{v}^{(0)}, n-1) \\ &\le \operatorname{codim}_{\mathbb{T}(\vec{0}, n)} \mathbb{T}(\vec{0}, n-1) \\ &= n + 1. \end{aligned}$$

□

Lemma 3.5 (Lipschitz continuity). *Let $p, q \in \mathcal{P}$ be distinct points. Suppose $\vec{v}^{(0)}, \vec{v}^{(1)}, \cdots \in \mathbb{Z}^{\mathcal{P}}$ are such that $\vec{v}^{(0)} \leq \vec{v}^{(1)} \leq \cdots$ coordinate-wise and strictly increasing at the coordinate indexed by p. Then*

$$0 \leq \sum_{r \geq 0} b_p(\vec{v}^{(r)}, n) - \sum_{r \geq 0} b_p(\vec{v}^{(r)} + \vec{e}_q, n) \leq n + 1.$$

Proof. Combine Lemmas 3.2, 3.3 and 3.4. □

3.4. **How the number of vanishing conditions varies with the handicap.** As in the previous subsection, let us continue to focus our attention on a set of points $\mathcal{P}$ on a fixed plane $F \cong \mathbb{R}^2$ (which we will drop from our notation temporarily).

Given a handicap $\vec{\alpha} \in \mathbb{Z}^{\mathcal{P}}$ (restricted to this plane), we define the vector $\vec{v}^{p,r}(\vec{\alpha})$ as follows. It assigns to coordinate $p' \in \mathcal{P}$ the smallest nonnegative integer r' such that $(p, r) \preceq (p', r')$. Equivalently, the value of $\vec{v}^{p,r}(\vec{\alpha})$ at p' is given by

$$v^{p,r}_{p'}(\vec{\alpha}) = \begin{cases} \max\{r - \alpha_p + \alpha_{p'} + 1, 0\} & \text{if } p' \text{ comes strictly before } p \text{ in the preassigned order,} \\ \max\{r - \alpha_p + \alpha_{p'}, 0\} & \text{otherwise.} \end{cases} \tag{3.3}$$

In other words, $\vec{v}^{p,r}(\vec{\alpha})$ collects the desired vanishing orders at each joint on F at the stage right before we hit (p, r) in the priority order.

Define $\mathcal{B}^r_p(\vec{\alpha}, n)$ and $\mathcal{B}_p(\vec{\alpha}, n)$ as in Section 3.2 restricted to this plane. Recall that for every $(p, r) \in \mathcal{P} \times \mathbb{Z}_{\geq 0}$, the disjoint union $\bigcup_{(p',r') \prec (p,r)} \mathcal{B}^{r'}_{p'}$ is basis of $\sum_{(p',r') \prec (p,r)} \mathbb{B}^{r'}_{p'}$. Then a polynomial $g \in \mathbb{R}[x_1, \ldots, x_6]_{\leq n}$ lies in the common kernel of $\bigcup_{(p',r') \prec (p,r)} \mathcal{B}^{r'}_{p'}$ if and only if the restriction of g to the plane F vanishes to order at least $v^{p,r}_q(\vec{\alpha})$ for every $q \in \mathcal{P}$. Since adding $\mathcal{B}^r_p$ makes this set a basis for $\sum_{(p',r') \preceq (p,r)} \mathbb{B}^{r'}_{p'}$, its size is the number of non-redundant constraints that we need to add to increase the order of vanishing at p by 1. Thus

$$\left|\mathcal{B}^r_p(\vec{\alpha}, n)\right| = b_p(\vec{v}, n) = \operatorname{codim}_{\mathbb{T}(\vec{v},n)} \mathbb{T}(\vec{v} + \vec{e}_p, n) \quad \text{ with } \vec{v} = \vec{v}^{p,r}(\vec{\alpha}). \tag{3.4}$$

The observations in the previous section then imply the following.

Lemma 3.6 (Bounded domain). *Let $n \geq 0$ and $\vec{\alpha} \in \mathbb{Z}^{\mathcal{P}}$. Let $p, q \in \mathcal{P}$. If $\alpha_p < \alpha_q - n$, then $|\mathcal{B}_p(\vec{\alpha}, n)| = 0$.*

Proof. For each $r \geq 0$, the value of $\vec{v} = \vec{v}^{p,r}(\vec{\alpha})$ at q is greater than n, so $\dim \mathbb{T}(\vec{v}, n) = 0$ by Lemma 3.1. Hence $\left|\mathcal{B}^r_p(\vec{\alpha}, n)\right| = b_p(\vec{v}, n) = 0$. □

Lemma 3.7 (Monotonicity). *Let n be a positive integer and $\vec{\alpha}^{(1)}, \vec{\alpha}^{(2)} \in \mathbb{Z}^{\mathcal{P}}$ be two handicaps. Suppose $p \in \mathcal{P}$ satisfies $\alpha^{(1)}_p - \alpha^{(1)}_{p'} \leq \alpha^{(2)}_p - \alpha^{(2)}_{p'}$ for all $p' \in \mathcal{P}$. Then $\left|\mathcal{B}_p(\vec{\alpha}^{(1)}, n)\right| \leq \left|\mathcal{B}_p(\vec{\alpha}^{(2)}, n)\right|$.*

Proof. For each $i = 1, 2$, let $\vec{v}^{(i)} = \vec{v}^{(p,r)}(\vec{\alpha}^{(i)})$. From (3.3) we see that $\vec{v}^{(1)} \geq \vec{v}^{(2)}$ coordinatewise and with equality at p. Then Lemma 3.2 gives $b_p(\vec{v}^{(1)}, n) \leq b_p(\vec{v}^{(2)}, n)$, and (3.4) gives the claim. □

Lemma 3.8 (Lipschitz continuity). *Let $p \in \mathcal{P}$ and $\vec{\alpha}^{(1)}, \vec{\alpha}^{(2)} \in \mathbb{Z}^{\mathcal{P}}$. Then*

$$\left| |\mathcal{B}_p(\vec{\alpha}^{(1)}, n)| - |\mathcal{B}_p(\vec{\alpha}^{(2)}, n)| \right| \leq (n + 1) \sum_{p' \in \mathcal{P}} \left| (\alpha^{(1)}_{p'} - \alpha^{(1)}_p) - (\alpha^{(2)}_{p'} - \alpha^{(2)}_p) \right|.$$

Proof. Shifting all handicaps by the same constant does not change the priority order and thus also does not change $|\mathcal{B}_p|$. Since the right-hand side of the above inequality is also invariant under translation we may assume that $\alpha^{(1)}_p = \alpha^{(2)}_p = 0$.

Starting with $\vec{\alpha} = \vec{\alpha}^{(1)}$, we can perform a sequence of changes where at each step we change the value of the handicap $\vec{\alpha}$ at some $p' \neq p$ by exactly 1, so that the vector $(\alpha_{p'})_{p' \in \mathcal{P}}$ ends up being equal to $(\alpha^{(2)}_{p'})$ after exactly $\sum_{p'\in\mathcal{P}}\left|\alpha^{(1)}_{p'}-\alpha^{(2)}_{p'}\right|$ moves. So it suffices to prove the inequality for each step in the process, i.e., showing that for every $\vec{\alpha}\in\mathbb{Z}^{\mathcal{P}}$ and $q\neq p$,

$$0\leq|\mathcal{B}_p(\vec{\alpha},n)|-|\mathcal{B}_p(\vec{\alpha}+\vec{e}_q,n)|\leq n+1.$$

The first inequality follows from Lemma 3.7. For the second inequality, by $|\mathcal{B}_p(\vec{\alpha},n)|=\sum_{r\geq 0}\left|\mathcal{B}^r_p(\vec{\alpha},n)\right|$ and (3.4), it suffices to prove

$$\sum_{r\geq 0}b_p(\vec{v}^{p,r}(\vec{\alpha}),n)-\sum_{r\geq 0}b_p(\vec{v}^{p,r}(\vec{\alpha}+\vec{e}_q),n)\leq n+1.$$

From (3.3), we see that there is some r_0 so that $\vec{v}^{p,r}(\vec{\alpha}+\vec{e}_q)=\vec{v}^{p,r}(\vec{\alpha})$ for all $r<r_0$ and $\vec{v}^{p,r}(\vec{\alpha}+\vec{e}_q)=\vec{v}^{p,r}(\vec{\alpha})+\vec{e}_q$ for all $r\geq r_0$. Restricting the sum to $r\geq r_0$ (the earlier terms cancel), we obtain the desired inequality by Lemma 3.5. ☐

3.5. **Vanishing lemma.** Now we start considering the interactions between different planes at the joints. The next statement is a vanishing lemma that is tailored to this joints problem. We omit the dependence on the handicap $\vec{\alpha}$ and the degree n from the notation since we are keeping them fixed in this subsection. Recall from the beginning of the section that, at each joint, we arbitrarily chose three planes that form this joint. Note that this vanishing lemma is the only place in the proof where we use the hypothesis that the three planes that form a joint do not all lie in some hyperplane.

Lemma 3.9. *Let $(\mathcal{J},\mathcal{F})$ be a joints configuration of planes in $\mathbb{R}^6$. Given a handicap $\vec{\alpha}\in\mathbb{Z}^{\mathcal{J}}$ and its associated priority order, and a positive integer n, choose $\mathcal{D}_{p,F}$ as earlier.*

Then for every nonzero polynomial $g\in\mathbb{R}[x_1,\ldots,x_6]$ of degree at most n, one has

$$D_1D_2D_3g(p)\neq 0$$

for some joint $p\in\mathcal{J}$ formed by $F_1,F_2,F_3\in\mathcal{F}$, and some $D_i\in\mathcal{D}_{p,F_i}$ for each $i=1,2,3$.

Proof. Suppose, on the contrary, that there were some nonzero $g\in\mathbb{R}[x_1,\ldots,x_6]_{\leq n}$ such that $D_1D_2D_3g(p)=0$ for every $p\in\mathcal{J}$, with $F_1,F_2,F_3\in\mathcal{F}$ being the three planes passing through p, and every $D_i\in\mathcal{D}_{p,F_i}$ for each $i=1,2,3$.

Choose $p\in\mathcal{J}$ to minimize $(p,v_p(g))$ under $\prec$, where $v_p(g)$ is the order of vanishing of g at p.

Recall that $\mathbb{D}^r_{p,F}$ is the space of r-th order derivative operators at p along F. Since g vanishes to order exactly $v_p(g)$ at p and the planes F_1,F_2,F_3 do not all line in one hyperplane, there exist $D_1\in\mathbb{D}^{r_1}_{p,F_1}$, $D_2\in\mathbb{D}^{r_2}_{p,F_2}$, $D_3\in\mathbb{D}^{r_3}_{p,F_1}$ with $D_1D_2D_3g(p)\neq 0$ and $r_1+r_2+r_3=v_p(g)$. Among all choices of D_1,D_2,D_3 (including choices of r_1,r_2,r_3), choose ones so that $|\{i\in[3]:D_i\in\mathcal{D}_{p,F_i}\}|$ is maximized. By the assumption at the beginning of the proof, one must have $D_i\notin\mathcal{D}_{p,F_i}$ for some i. Relabeling if necessary, assume that $D_1\notin\mathcal{D}_{p,F_1}$.

Suppose $p'\in F_1\cap\mathcal{J}$ and $r'\in\mathbb{Z}_{\geq 0}$ satisfy $(p',r')\prec(p,r_1)$. We get $(p',r'+r_2+r_3)\prec(p,r_1+r_2+r_3)=(p,v_p(g))$. By the choice of p, we have $(p,v_p(g))\preceq(p',v_{p'}(g))$. Thus $(p',r'+r_2+r_3)\prec(p',v_{p'}(g))$, and hence $r'+r_2+r_3<v_{p'}(g)$. If follows that $DD_2D_3g(p')=0$ for all $D\in\mathbb{D}^{r'}_{p',F_1}$ by the definition of vanishing order.

From the above paragraph we deduce that D_2D_3g lies in the common kernel of $\mathbb{B}^{r'}_{p',F_1}$ ranging over all $(p',r')\in(F_1\cap\mathcal{J})\times\mathbb{Z}_{\geq 0}$ with $(p',r')\prec(p,r_1)$. Since $D_1D_2D_3g(p)\neq 0$, we deduce that D_2D_3g does not lie in the common kernel of $\mathcal{B}^{r_1}_{p,F_1}$, i.e., there is some $D\in\mathcal{D}^{r_1}_{p,F_1}$ with $DD_2D_3g(p)\neq 0$. But this D contradicts the earlier assumption that the choice of (D_1,D_2,D_3) maximizes $|\{i:D_i\in\mathcal{D}_{p,F_i}\}|$. ☐

The next inequality uses parameter counting.

Lemma 3.10. *Assume the same setup as Lemma 3.9. We have*

$$\sum_{p\in\mathcal{J}}\prod_{F\ni p}|\mathcal{D}_{p,F}(\vec{\alpha},n)|\geq\binom{n+6}{6}.$$

Proof. Denote the left hand side by A and right hand side by B. Consider the constraints on $g \in \mathbb{R}[x_1, \dots, x_6]_{\le n}$ where for all $p \in \mathcal{J}$ formed by the planes $F_1, F_2, F_3 \in \mathcal{F}$, we require

$$D_1 D_2 D_3 g(p) = 0 \quad \forall D_i \in \mathcal{D}_{p,F_i}, i = 1, 2, 3.$$

This requirement is asking A linear functionals on $\mathbb{R}[x_1, \dots, x_6]_{\le n}$, which has dimension B, to vanish at g. Hence, if $A < B$, then there exists a nonzero polynomial g in $\mathbb{R}[x_1, \dots, x_6]_{\le n}$ that satisfies all the conditions, which would contradict Lemma 3.9. □

3.6. **Choosing the handicaps.** We say that a joints configuration $(\mathcal{J}, \mathcal{F})$ is *connected* if the following graph is connected: the vertex set is $\mathcal{J}$, with two joints adjacent if there is some plane in $\mathcal{F}$ containing both joints.

Lemma 3.11. *Let $(\mathcal{J}, \mathcal{F})$ be any connected joints configuration, and let n be some positive integer. Then there exists a choice of handicap $\vec{\alpha} \in \mathbb{Z}^{\mathcal{J}}$ such that*

$$\max_{p \in \mathcal{J}} \prod_{F \ni p} \frac{|\mathcal{D}_{p,F}(\vec{\alpha}, n)|}{\binom{n+2}{2}} - \min_{p \in \mathcal{J}} \prod_{F \ni p} \frac{|\mathcal{D}_{p,F}(\vec{\alpha}, n)|}{\binom{n+2}{2}} \le \frac{C}{n}$$

for some constant C that only depends on $(\mathcal{J}, \mathcal{F})$ but not n.

Proof. Fix n throughout the proof. Denote

$$W_p(\vec{\alpha}) = \prod_{F \ni p} \frac{|\mathcal{D}_{p,F}(\vec{\alpha}, n)|}{\binom{n+2}{2}}$$

for all $p \in \mathcal{J}$. The α_p are arbitrary integers. However, note that shifting all α_p by the same constant does not affect the priority order and thus does not affect $W_p(\alpha)$. Furthermore, by Lemma 3.6, if two handicaps differ by more than n at two points on the same plane, then $W_p(\vec{\alpha}) = 0$. Therefore, there are only finitely many possibilities for the vector $(W_p(\vec{\alpha}) : p \in \mathcal{J})$. Among those possibilities, choose the one so that after sorting $W_p(\vec{\alpha})$ in descending order, this vector is least in lexicographical order over all such possible vectors. Suppose that the sorted result is

$$W_{p_1}(\vec{\alpha}) \ge W_{p_2}(\vec{\alpha}) \ge \cdots \ge W_{p_{|\mathcal{J}|}}(\vec{\alpha}).$$

We will show that $W_{p_i}(\vec{\alpha}) - W_{p_{i+1}}(\vec{\alpha}) \le C'/n$ for some constant C' to be determined. This will imply the desired statement.

Suppose for the sake of contradiction that the above claim does not hold. Let t be the least positive integer such that $W_{p_t}(\vec{\alpha}) - W_{p_{t+1}}(\vec{\alpha}) > C'/n$. Then let $\vec{v} = \vec{e}_{p_1} + \cdots + \vec{e}_{p_t}$ and let $\vec{\alpha}' = \vec{\alpha} - \vec{v}$ be a new handicap. We will consider the difference between $W_p(\vec{\alpha})$ and $W_p(\vec{\alpha}')$. By Lemma 3.8,

$$\left| |\mathcal{D}_{p,F}(\vec{\alpha}, n)| - |\mathcal{D}_{p,F}(\vec{\alpha}', n)| \right| \le |\mathcal{J}| \, (n+1) \le \frac{2\,|\mathcal{J}|}{n} \binom{n+2}{2}$$

for each joint p on each plane F. We have $|\mathcal{D}_{p,F}(\vec{\alpha}, n)| \le \binom{n+2}{2}$ by (3.1). We use the following telescoping inequality. For $x_1, x_2, x_3, y_1, y_2, y_3 \in [0, 1]$,

$$|x_1x_2x_3 - y_1y_2y_3| \le |x_1 - y_1| x_2x_3 + |x_2 - y_2| y_1x_3 + |x_3 - y_3| y_1y_2 \le 3 \max_i |x_i - y_i|.$$

Thus

$$\left| W_p(\vec{\alpha}') - W_p(\vec{\alpha}) \right| \le \frac{6|\mathcal{J}|}{n} = \frac{C'}{2n}$$

where we choose $C' = 12|\mathcal{J}|$.

By the monotonicity established in Lemma 3.7, we know that $W_{p_i}(\vec{\alpha}') \le W_{p_i}(\vec{\alpha})$ for $i \le t$, and $W_{p_i}(\vec{\alpha}') \ge W_{p_i}(\vec{\alpha})$ for $i > t$. By (3.1), we know that if $W_p(\vec{\alpha}') \ne W_p(\vec{\alpha})$ for some p, then there exists $i \le t$ such that $W_{p_i}(\vec{\alpha}') < W_{p_i}(\vec{\alpha})$. However, since the difference between $W_p(\vec{\alpha})$ and $W_p(\vec{\alpha}')$ is at most $C'/2n$, and $W_{p_t}(\vec{\alpha}) - W_{p_{t+1}}(\vec{\alpha}) > C'/n$, we know that $W_{p_1}(\vec{\alpha}'), \dots, W_{p_t}(\vec{\alpha}')$ are still the t

largest values among $(W_p(\vec{\alpha}'))_{p\in\mathcal{J}}$. This shows that $\vec{\alpha}'$ gives a strictly lower lexicographical order of $(W_p(\vec{\alpha}'))_{p\in\mathcal{J}}$, which is a contradiction.

Hence $W_p(\vec{\alpha}) = W_p(\vec{\alpha}') = W_p(\vec{\alpha}-\vec{v})$ for all $p \in \mathcal{J}$. By the same argument, we know that $W_p(\vec{\alpha}) = W_p(\vec{\alpha}') = W_p(\vec{\alpha}-c\vec{v})$ holds for $p \in \mathcal{J}$ for any positive integer c. By connectedness, we can find some $i \le t < j$ such that p_i and p_j are on the same plane. As a consequence, if c is chosen sufficiently large such that $\alpha_{p_i} - c < \alpha_{p_j} - n$, this implies that $W_{p_i}(\vec{\alpha}-c\vec{v}) = 0$. By our ordering this implies that $W_{p_{i'}}(\vec{\alpha}-c\vec{v}) = 0$ for all $i' \ge i$. In particular, $W_{p_t}(\vec{\alpha}) = W_{p_{t+1}}(\vec{\alpha}) = 0$, contradicting our earlier assumption that $W_{p_t}(\vec{\alpha}) - W_{p_{t+1}}(\vec{\alpha}) > C'/n$. □

We are now ready to prove the joints theorem for a set of planes in $\mathbb{R}^6$.

Proof that N planes in $\mathbb{R}^6$ have $\sqrt{10/3}N^{3/2}$ joints. Assume first that the joints configuration is connected. Let n be some large positive integer. In this proof we will use O-notation to suppress constants that can depend on $(\mathcal{J},\mathcal{F})$ arbitrarily as long as they are independent of n. Choose $\vec{\alpha}$ according to Lemma 3.11. Then there exists W such that

$$\left| \prod_{F\ni p} \frac{|\mathcal{D}_{p,F}|}{\binom{n+2}{2}} - W \right| \le \frac{C}{n}$$

for all $p \in \mathcal{J}$. By Lemma 3.10, we have

$$|\mathcal{J}| W \binom{n+2}{2}^3 \ge \binom{n+6}{6} - O(n^5).$$

Therefore

$$W \ge \frac{8}{6! \cdot |\mathcal{J}|} - O(n^{-1}).$$

So there is some constant $c > 0$ (depending on $\mathcal{J}$ but not on n) so that $W \in [c, 1]$ for all sufficiently large n. For each $p \in \mathcal{J}$, by a Taylor series approximation,

$$W^{1/3} = \left(\prod_{F\ni p} \frac{|\mathcal{D}_{p,F}|}{\binom{n+2}{2}} \right)^{1/3} + O(n^{-1}).$$

Hence (in the summations, p ranges over joints and F ranges over planes in $\mathcal{F}$),

$$\begin{aligned}
3|\mathcal{J}| \left(\frac{W}{N^3}\right)^{1/3} &= 3\sum_p \left(\prod_{F\ni p} \frac{|\mathcal{D}_{p,F}|}{N\binom{n+2}{2}} \right)^{1/3} + O(n^{-1}) \\
&\le \sum_p \sum_{F\ni p} \frac{|\mathcal{D}_{p,F}|}{N\binom{n+2}{2}} + O(n^{-1}) && \text{[by AM-GM]} \\
&= \sum_F \sum_{p\in F} \frac{|\mathcal{D}_{p,F}|}{N\binom{n+2}{2}} + O(n^{-1}) \\
&= \sum_F \frac{1}{N} + O(n^{-1}) && \text{[by (3.1)]} \\
&= 1 + O(n^{-1}).
\end{aligned}$$

Thus

$$W \le \frac{N^3}{27|\mathcal{J}|^3} + O(n^{-1}).$$

By comparing the leading term in the upper bound and the lower bound of W, i.e., letting n go to infinity, we get that

$$\frac{8}{6! \cdot |\mathcal{J}|} \leq \frac{N^3}{27|\mathcal{J}|^3},$$

and by rearranging we get that

$$|\mathcal{J}| \leq \sqrt{\frac{10}{3}} N^{3/2}.$$

The above argument proves the result for connected joints configurations. In general, decompose the joints configuration $(\mathcal{J}, \mathcal{F})$ into connected components (in the sense of the associated graph) $(\mathcal{J}_1, \mathcal{F}_1), \ldots, (\mathcal{J}_k, \mathcal{F}_k)$. Denote $N_i = |\mathcal{F}_i|$. Then

$$|\mathcal{J}| = \sum_{i=1}^{k} |\mathcal{J}_i| \leq \sqrt{\frac{10}{3}} \sum_{i=1}^{k} N_i^{3/2} \leq \sqrt{\frac{10}{3}} N^{3/2}. \qquad \square$$

Remark. The arguments here generalize straightforwardly to joints of flats in arbitrary dimensions.

4. Derivatives along varieties

In this section we discuss how to generalize the argument in Section 3 to varieties in $\mathbb{F}^d$. There are two issues that we need to address. The first is to define appropriate higher order directional derivatives along varieties. As we explain below, it does not suffice to simply take derivatives along the tangent plane, as those miss the higher order data of the variety. The second is to generalize derivatives from the reals to general fields. Since we are working with polynomials, differentiation can be viewed as a formal algebraic operation. To handle fields of positive characteristics, we use Hasse derivatives.

Let V be a k-dimensional variety in $\mathbb{F}^n$. Let $I(V)$ be the ideal of polynomials in $\mathbb{F}[x_1, \ldots, x_d]$ that vanish on V. Define $R_V = \mathbb{F}[x_1, \ldots, x_d]/I(V)$. The elements of R_V are called *regular functions* on V. Let p be a regular point on V, that is, a point where the Zariski tangent space of V at p is also k-dimensional. Given a nonnegative integer r, we would like to write down derivative operators D on $\mathbb{F}[x_1, \ldots, x_d]$ so that $Dg(p)$ is well defined not just when $g \in \mathbb{F}[x_1, \ldots, x_d]$, but also when g is a regular function on V. The point here is that regular functions on V may be represented as polynomials in $\mathbb{F}[x_1, \ldots, x_d]$ in non-unique ways (by adding a polynomial that vanishes on V), but we should study derivative operators D whose evaluation $Dg(p)$ does not depend on this representation of g.

4.1. An explicit example.

We consider the explicit example of the circle V in $\mathbb{R}^2$ centered at $(0, 1/2)$ of radius $1/2$. In particular, V is defined by the equation $y = x^2 + y^2$. Let $p = (0, 0)$ be the origin. How should we define a second-order derivative at p along V?

Naively one might take $\partial^2/\partial x^2$ since the tangent at p is the x-coordinate direction. However, consider evaluation of this derivative at p applied to the two sides of $y = x^2 + y^2$ (an identity of regular functions on V): the left-hand side gives 0 while the right-hand side gives 2. So $\partial^2/\partial x^2$ does not induce a linear functional on the space of regular functions on V.

To fix this issue, we can rewrite all regular functions on V as power series centered at p using the *local coordinate* x of V. Indeed, by repeated substituting $y \leftarrow x^2 + y^2$, we can write y as a power series in x:

$$\begin{aligned} y &= x^2 + y^2 \\ &= x^2 + (x^2 + y^2)^2 \\ &= x^2 + (x^2 + (x^2 + y^2)^2)^2 \\ &= x^2 + x^4 + 2x^6 + \cdots \end{aligned}$$

We would like a derivative operator D on $\mathbb{R}[x,y]$ so that $Dg(0,0)$ equals to the coefficient of x^2 in $g(x, x^2+x^4+2x^6+\cdots)$, which in turn equals to the coefficient of x^2 plus the coefficient of y in $g(x,y)$. It is not hard to see that only such choice is $\frac{1}{2}\frac{\partial^2}{\partial x^2}+\frac{\partial}{\partial y}$. Conversely, it is not hard to check that $Dg(0,0)=0$ for every $g\in\mathbb{R}[x,y]$ that vanishes identically on V.

Elaborating on this example further, for each nonnegative integer r, we will define $\mathbb{D}^r_{p,V}$ to be a one-dimensional space spanned a derivative operator D on $\mathbb{R}[x,y]$ such that $Dg(0,0)$ equals to the coefficient of x^r in $g(x,x^2+x^4+2x^6+\cdots)$. Thus (here $\langle\cdot\rangle$ denotes the span)

$$\begin{aligned}
\mathbb{D}^0_{p,V} &= \langle \mathrm{Id}\rangle \\
\mathbb{D}^1_{p,V} &= \left\langle \frac{\partial}{\partial x}\right\rangle \\
\mathbb{D}^2_{p,V} &= \left\langle \frac{1}{2}\frac{\partial^2}{\partial x^2}+\frac{\partial}{\partial y}\right\rangle \\
\mathbb{D}^3_{p,V} &= \left\langle \frac{1}{6}\frac{\partial^3}{\partial x^3}+\frac{\partial^2}{\partial x\partial y}\right\rangle \\
\mathbb{D}^4_{p,V} &= \left\langle \frac{1}{24}\frac{\partial^4}{\partial x^4}+\frac{1}{2}\frac{\partial^3}{\partial x^2\partial y}+\frac{1}{2}\frac{\partial^2}{\partial y^2}\right\rangle \\
&\ \ \vdots
\end{aligned}$$

Then, for each each $D\in\mathbb{D}^r_{p,V}$, the map sending $g\in\mathbb{R}[x,y]$ to $Dg(0,0)$ passes to a linear functional on the space $R_V=\mathbb{R}[x,y]/I(V)$ of regular functions on V.

The computation in the above example can be extended to any variety over any field, as we explain below.

4.2. **Local coordinates.** Given a regular point p on a k-dimensional variety V, after a translation and a linear change of coordinates, suppose that p is at the origin and the first k coordinate vectors are tangent to V. Then by assumption, there are polynomials $f_{k+1},\dots,f_d$ without any constant or linear terms so that on V, we have $x_{k+1}=f_{k+1}(x_1,\dots,x_k)$, $\dots$, $x_d=f_d(x_1,\dots,x_k)$. For each $i=k+1,\dots,d$, by repeated substitutions using the defining equations, as functions on V, we can write each x_i as a formal power series $h_i(x_1,\dots,x_k)$ in the *local coordinates* $x_1,\dots,x_k$ for V at p.

The procedure of taking a power series described earlier can be described in algebraic geometry as a *completion.* We give a quick summary here and refer the reader to a standard algebraic geometry textbook, e.g., [11, Chapter 7] [31, Chapter 29]. Let p be a regular point on a k-dimensional variety V in $\mathbb{F}^d$. Let $\mathfrak{m}_p\subset R_V$ be the maximal ideal of regular functions that vanish at p. Then the *completion* $\widehat{R_{p,V}}$ of R_V at p is the inverse limit $\varprojlim R_V/\mathfrak{m}_p^m$. The family of projection maps $R_V\to R_V/\mathfrak{m}_p^m$ induces a map $\iota_{p,V}: R_V\to\widehat{R_{p,V}}$.

The completion should be thought of as the ring of formal power series around p. For example, when $R_V=\mathbb{F}[x]$ and $\mathfrak{m}_p=(x)$, the completion is the ring of formal power series $\mathbb{F}\llbracket x\rrbracket$. More generally, for a regular point p on V, assuming that p is the origin and $x_1,\dots,x_k\in\mathfrak{m}_p$ span the Zariski cotangent space $\mathfrak{m}_p/\mathfrak{m}_p^2$, the map $\mathbb{F}\llbracket x_1,\dots,x_k\rrbracket\to\widehat{R_{p,v}}$ sending x_i to $\iota_{p,V}(x_i)$ is an isomorphism (say, by the Cohen structure theorem). In other words, there is a *local coordinate system* at p so that every regular function on V can be written as a formal power series around p.

It will be useful to know that the formal power series expansion of a regular function is zero if and only if the regular function is zero, i.e., the completion map $R_V\to\widehat{R_{p,V}}$ is injective. This fact follows from the Krull intersection theorem below (recall that our varieties are always irreducible).

Theorem 4.1 (Krull intersection theorem)**.** *Let R be an integral domain and I be a proper ideal of R. Then $\bigcap_{m=0}^\infty I^m=\{0\}$.*

4.3. **Hasse derivatives.** In the explicit example earlier, the main goal of taking derivatives is to extract coefficients. This is a formal algebraic procedure that does not rely on real analysis. To allow for arbitrary fields, including those of positive characteristics, we use an algebraic variant known as Hasse derivatives, whose definition and basic properties we summarize below. For proofs of these basic properties of Hasse derivatives, we refer the reader to [10], where Hasse derivatives were used to study the finite field Kakeya problem.

Definition 4.2 (Hasse derivatives)**.** For any d-tuple $\vec{\omega} = (\omega_1, \dots, \omega_d)$ of nonnegative integers, define $\mathsf{H}^{\vec{\omega}}$ to be the linear operator on $\mathbb{F}[x_1, \dots, x_d]$ given by (writing $x^\delta = x_1^{\delta_1} \cdots x_d^{\delta_d}$ and $\binom{\vec{\gamma}}{\vec{\omega}} := \binom{\gamma_1}{\omega_1} \cdots \binom{\gamma_d}{\omega_d}$)

$$\mathsf{H}^{\vec{\omega}} x^{\vec{\delta}} = \binom{\vec{\delta}}{\vec{\omega}} x^{\vec{\delta} - \vec{\omega}}$$

for every d-tuple $\vec{\delta} = (\delta_1, \dots, \delta_d)$ of nonnegative integers.

In particular, $\mathsf{H}^{\vec{\omega}} x^{\vec{\delta}} = 0$ unless $\vec{\delta} \geq \vec{\omega}$ coordinatewise.

Over the reals, it is not hard to see that the two notions of derivatives are related by a constant factor

$$\mathsf{H}^{\vec{\omega}} = \frac{1}{\vec{\omega}!} \frac{\partial^{\vec{\omega}}}{\partial x^{\vec{\omega}}} := \frac{1}{\omega_1! \cdots \omega_d!} \frac{\partial^{\omega_1 + \cdots + \omega_d}}{\partial x_1^{\omega_1} \cdots \partial x_d^{\omega_d}}.$$

Like usual derivatives, Hasse derivatives commute:

$$\mathsf{H}^{\vec{\alpha}} \mathsf{H}^{\vec{\beta}} = \binom{\vec{\alpha} + \vec{\beta}}{\vec{\alpha}} \mathsf{H}^{\vec{\alpha} + \vec{\beta}} = \mathsf{H}^{\vec{\beta}} \mathsf{H}^{\vec{\alpha}}.$$

Hasse derivatives form an algebraic generalization of the usual derivatives when acting on polynomials or formal power series. The evaluation of a Hasse derivative corresponds to coefficient extraction (without the factorial factors that might be troublesome in fields of positive characteristics). Indeed, we have the following "Taylor's theorem": given formal variables $x_1, \dots, x_d, y_1, \dots, y_d$ and a polynomial $g \in \mathbb{F}[x_1, \dots, x_d]$, we have

$$g(x + y) = \sum_{\vec{\omega} \in \mathbb{Z}_{\geq 0}^d} (\mathsf{H}^{\vec{\omega}} g)(x) y^{\vec{\omega}} \tag{4.1}$$

for any $g \in \mathbb{F}[x_1, \dots, x_d]$. This identity can be easily checked for each monomial $g(x) = x^{\vec{\delta}}$. From this characterization, we see that Hasse derivatives behave well under affine coordinate transforms (as we would expect for derivatives). For example, it makes sense to talk about directional Hasse derivatives without specifying a choice of a coordinate system.

4.4. **Higher order directional derivatives.** Now that we have the tools of completion and Hasse derivatives, we are ready to define higher order directional derivatives at a regular point p along a k-dimensional variety V in $\mathbb{F}^d$, generalizing the notion for flats from Section 3.

By an affine change of coordinates, assume that p is at the origin, and the tangent space of V at p is spanned by the first k coordinate directions. For each $i = k+1, \dots, d$, write each x_i as a formal power series $h_i(x_1, \dots, x_k)$ in the "local coordinates" $x_1, \dots, x_k$ for V at p. Equivalently, $h_i(x_1, \dots, x_k)$ is the image of x_i under the completion map $R_V \to \widehat{R_{p,V}} \cong \mathbb{F}[\![x_1, \dots, x_k]\!]$.

We define $\mathbb{D}_{p,V}^r$ to be the space of all linear combinations D of Hasse derivative operators on $\mathbb{F}[x_1, \dots, x_d]$ such that the map $\mathbb{F}[x_1, \dots, x_d] \to \mathbb{F}$ defined by $g \mapsto Dg(p)$ equals a linear form on coefficients of the homogeneous degree r part of

$$\widehat{g}(x_1, \dots, x_k) := g(x_1, x_2, \dots, x_k, h_{k+1}(x_1, \dots, x_k), \dots, h_d(x_1, \dots, x_k)),$$

which is the power series representation of g as a regular function on V in local coordinates at p. Let us also write out this definition more explicitly. Given $(\gamma_1, \ldots, \gamma_k) \in \mathbb{Z}_{\geq 0}^k$, define

$$D_{p,V}^{\vec{\gamma}} = \sum_{\vec{\omega} \in \mathbb{Z}_{\geq 0}^d} c_{\vec{\omega}}^{\vec{\gamma}} \mathsf{H}^{\vec{\omega}}$$

where

$$c_{\vec{\omega}}^{\vec{\gamma}} = \text{the coefficient of } x_1^{\gamma_1} \cdots x_k^{\gamma_k} \text{ in } x_1^{\omega_1} \cdots x_k^{\omega_k} h_{k+1}(x_1, \ldots, x_k)^{\omega_{k+1}} \cdots h_d(x_1, \ldots, x_k)^{\omega_d}$$

Then $D_{p,V}^{\vec{\gamma}} g(0, \ldots, 0)$ equals the coefficient of $x^{\vec{\gamma}}$ in $\widehat{g}(x_1, x_2, \ldots, x_k)$. We then set

$$\mathbb{D}_{p,V}^r = \operatorname{span}\left\{ D_{p,V}^{\vec{\gamma}} : \vec{\gamma} = (\gamma_1, \ldots, \gamma_k) \in \mathbb{Z}_{\geq 0}^k, \gamma_1 + \cdots + \gamma_k = r \right\}.$$

Note that the $D_{p,V}^{\vec{\gamma}}$ in the above set are linearly independent. To see this, first note that because no h_i has constant or linear terms, one has

$$D_{p,V}^{\vec{\gamma}} \in \mathsf{H}^{(\vec{\gamma}, 0, \ldots, 0)} + \operatorname{span}\left\{ \mathsf{H}^{\vec{\omega}} : \omega_1 + \cdots + \omega_d < \gamma_1 + \cdots + \gamma_k \right\}. \tag{4.2}$$

The Hasse derivative operators $\mathsf{H}^{\vec{\omega}}$ are linearly independent as $\vec{\omega}$ ranges over $\mathbb{Z}_{\geq 0}^d$. Since the top weight component of $D_{p,V}^{\vec{\gamma}}$ is $\mathsf{H}^{(\vec{\gamma}, 0, \ldots, 0)}$, we see that the $D_{p,V}^{\vec{\gamma}}$'s are linearly independent as $\vec{\gamma}$ ranges over $\mathbb{Z}_{\geq 0}^k$.

The key property, as well as the motivation for the above definition, is that for every $D \in \mathbb{D}_{p,V}^r$, there is a well defined map $R_V \to \mathbb{F}$ given by $g \mapsto Dg(p)$. To define this derivative evaluation, we can replace $g \in R_V$ by a representative $g \in \mathbb{F}[x_1, \ldots, x_d]$, and we need to check that $Dg(p)$ does not depend on the choice of the representative. Indeed, if g is identically zero on V, then $\widehat{g} = 0$, and hence $Dg(p) = 0$.

The above explicit formula defines $\mathbb{D}_{p,V}^r$ assuming that p is at the origin and the tangent space of V at p is spanned by the first k coordinate directions. By an affine transformation (using (4.1) to determine the behavior of Hasse derivatives under affine transformations), we can define the space $\mathbb{D}_{p,V}^r$ of r-th order directional derivatives at any regular point p on a variety V.

Having defined $\mathbb{D}_{p,V}^r$, we now can proceed nearly identically as in Section 3 to prove the joints theorem for varieties. Details are given in the next section.

5. Proof of the main theorem

5.1. **Priority order, handicaps, and a choice of basis.** Given a set of joints $\mathcal{J}$ with a fixed preassigned order, and a handicap $\vec{\alpha} \in \mathbb{Z}^{\mathcal{J}}$, we define the priority order $\prec$ on $\mathcal{J} \times \mathbb{Z}_{\geq 0}$ as before.

Let n be a positive integer. Let $R_{V,\leq n}$ denote the space of regular functions on V that can be represented as a polynomial of degree at most n in $x_1, \ldots, x_n$. In other words, $R_{V,\leq n}$ is the image of $\mathbb{F}[x_1, \ldots, x_d]_{\leq n}$ under the projection $\mathbb{F}[x_1, \ldots, x_d] \to R_V$.

Define $\mathbb{B}_{p,V}^r(n)$ to be the set of linear functionals on $R_{V,\leq n}$ of the form $g \mapsto Dg(p)$ for some $D \in \mathbb{D}_{p,V}^r$ (this is a well defined linear functional as explained earlier). Note that $g \in R_{V,\leq n}$ vanishes under $\mathbb{B}_{p,V}^0(n) + \cdots + \mathbb{B}_{p,V}^{r-1}(n)$ if and only if g vanishes to order at least r at p. Here a regular function g on V *vanishes at p to order at least r* if $g \in \mathfrak{m}_{p,V}^r$ where $\mathfrak{m}_{p,V}$ is the maximal ideal of R_V corresponding to p. Equivalently, power series representation of g using local coordinates at p has no terms with degree lower than r.

Now, exactly as in Section 3.2, we go through all pairs $(p, r) \in (\mathcal{J} \cap V) \times \mathbb{Z}_{\geq 0}$ according to the priority order and choose sets $\mathcal{B}_{p,V}^r(\vec{\alpha}, n) \subset \mathbb{B}_{p,V}^r(n)$ as earlier so that the disjoint union $\bigcup_{(p',r') \preceq (p,r)} \mathcal{B}_{p',V}^{r'}(\vec{\alpha}, n)$ is a basis of $\sum_{(p',r') \preceq (p,r)} \mathbb{B}_{p',V}^{r'}(\vec{\alpha}, n)$. Choose $\mathcal{D}_{p,V}^r(\vec{\alpha}, n) \subset \mathbb{D}_{p,V}^r$ with the same size as $\mathcal{B}_{p,V}^r(\vec{\alpha}, n)$ so that $\mathcal{B}_{p,V}^r(\vec{\alpha}, n) = \{g \mapsto Dg(p) : D \in \mathcal{D}_{p,V}^r(\vec{\alpha}, n)\}$. Finally, write $\mathcal{B}_{p,V}(\vec{\alpha}, n) := \bigcup_{r \geq 0} \mathcal{B}_{p,V}^r(\vec{\alpha}, n)$ and $\mathcal{D}_{p,V}(\vec{\alpha}, n) := \bigcup_{r \geq 0} \mathcal{D}_{p,V}^r(\vec{\alpha}, n)$.

From the Krull intersection theorem, it follows that for every $p \in V$, $\sum_{r\geq 0} \mathbb{B}^r_{p,V}(\vec{\alpha}, n)$ spans the dual space of $R_{V,\leq n}$. Hence the disjoint union $\bigcup_{p\in V} \mathcal{B}_{p,V}(\vec{\alpha}, n)$ is a basis of the space of linear forms on $R_{V,\leq n}$. Thus

$$\sum_{p\in\mathcal{J}\cap V} |\mathcal{B}_{p,V}(\vec{\alpha}, n)| = \dim R_{V,\leq n} = \deg V \binom{n}{\dim V} + O_V(n^{\dim V - 1}). \tag{5.1}$$

Furthermore there is some $n_0(V)$ so that $\dim R_{V,\leq n}$ is a polynomial in n for all $n \geq n_0(V)$. This is a standard fact about the Hilbert series for a variety (see, e.g., [31, Chapter 18.6]).

5.2. **Regular functions with given vanishing orders.** This subsection parallels Section 3.3. Here we fix a k-dimensional variety V and a finite set of points $\mathcal{P} \subset V$. Given a vector $\vec{v} \in \mathbb{Z}^{\mathcal{P}}_{\geq 0}$, define

$$\mathbb{T}(\vec{v}, n) = \{g \in R_{V,\leq n} : g \text{ vanishes to order} \geq v_p \text{ at each } p \in \mathcal{P}\}.$$

Set $b_p(\vec{v}, n) := \operatorname{codim}_{\mathbb{T}(\vec{v},n)} \mathbb{T}(\vec{v} + \vec{e}_p, n)$.

Lemma 5.1 (Bounded domain)**.** *For every n there is some $C_V(n)$ so that if $\vec{v} \in \mathbb{Z}^{\mathcal{P}}_{\geq 0}$ has $\max_{p\in\mathcal{P}} v_p > C_V(n)$ then $\dim \mathbb{T}(\vec{v}, n) = 0$.*

Proof. By the Krull intersection theorem, $\bigcap_{m\geq 0} \mathfrak{m}^m_{p,V} = \{0\}$. Since $R_{V,\leq n}$ is finite dimensional, there exists $C = C_V(n)$ such that $\mathfrak{m}^C_{p,V} \cap R_{V,\leq n} = \{0\}$. Hence, if $\vec{v} \in \mathbb{Z}^{\mathcal{J}\cap V}_{\geq 0}$ satisfies $v_p \geq C$, then $\mathbb{T}_V(\vec{v}, n) = \{0\}$. □

We omit the proofs of the next two lemmas, which mirror those of Section 3.3, except to note that the last line of the proof of Lemma 3.4 should be adapted as

$$\operatorname{codim}_{\mathbb{T}(\vec{0},n)} \mathbb{T}(\vec{0}, n-1) = \dim R_{V,\leq n} - \dim R_{V,\leq n-1} = \deg V \binom{n}{\dim V - 1} + O_V(n^{\dim V - 2}).$$

To see this we use the fact that $\dim R_{V,\leq n}$, for sufficiently large n, equals to a polynomial (the Hilbert polynomial) whose leading term given in (5.1). The right-hand side is the finite difference of this polynomial which can readily be seen to have the above form.

Lemma 5.2 (Monotonicity)**.** *Let $p \in \mathcal{P}$. Suppose $\vec{v}^{(1)}, \vec{v}^{(2)} \in \mathbb{Z}^{\mathcal{P}}_{\geq 0}$ satisfy $\vec{v}^{(1)} \geq \vec{v}^{(2)}$ coordinatewise and with equality at p. Then $b_p(\vec{v}^{(1)}, n) \leq b_p(\vec{v}^{(2)}, n)$ for all n.*

Lemma 5.3 (Lipschitz continuity)**.** *Let $p, q \in \mathcal{P}$ be distinct points. Suppose $\vec{v}^{(0)}, \vec{v}^{(1)}, \dots \in \mathbb{Z}^{\mathcal{P}}$ are such that $\vec{v}^{(0)} \leq \vec{v}^{(1)} \leq \cdots$ coordinate-wise and strictly increasing at the coordinate indexed by p. Then*

$$0 \leq \sum_{r\geq 0} b_p(\vec{v}^{(r)}_p, n) - \sum_{r\geq 0} b_p(\vec{v}^{(r)}_p + \vec{e}_q, n) \leq \deg V \binom{n}{\dim V - 1} + O_V(n^{\dim V - 2}).$$

5.3. **How the number of vanishing conditions varies with the handicap.** The lemmas in Section 3.4 can now be easily adapted to varieties. As in the previous subsection, we continue to focus our attention on a set of points $\mathcal{P}$ on a variety V.

Given a handicap $\vec{\alpha} \in \mathbb{Z}^{\mathcal{P}}$ (restricted to V), we define the vector $\vec{v}^{p,r}(\vec{\alpha})$ identically to Section 5.3. We have

$$\left|\mathcal{B}^r_{p,V}(\vec{\alpha}, n)\right| = b_p(\vec{v}, n) = \operatorname{codim}_{\mathbb{T}(\vec{v},n)} \mathbb{T}(\vec{v} + \vec{e}_p, n) \quad \text{ with } \vec{v} = \vec{v}^{p,r}(\vec{\alpha}).$$

We omit the proofs of the following lemmas, which mirror those of Section 3.4 but now using the lemmas from the previous subsection.

Lemma 5.4 (Bounded domain)**.** *For each n there is some $C_V(n)$ so that if $p \in \mathcal{P}$ and $\vec{\alpha} \in \mathbb{Z}^{\mathcal{P}}$ satisfy $\alpha_p < \max_{q\in\mathcal{P}} \alpha_q - C_V(n)$, then $|\mathcal{B}_{p,V}(\vec{\alpha}, n)| = 0$.*

Lemma 5.5 (Monotonicity)**.** *Let n be a positive integer and $\vec{\alpha}^{(1)}, \vec{\alpha}^{(2)} \in \mathbb{Z}^{\mathcal{P}}$ be two handicaps. Suppose $p \in \mathcal{P}$ satisfies $\alpha_p^{(1)} - \alpha_{p'}^{(1)} \leq \alpha_p^{(2)} - \alpha_{p'}^{(2)}$ for all $p' \in \mathcal{P}$. Then $\left|\mathcal{B}_{p,V}(\vec{\alpha}^{(1)}, n)\right| \leq \left|\mathcal{B}_{p,V}(\vec{\alpha}^{(2)}, n)\right|$.*

Lemma 5.6 (Lipschitz continuity)**.** *Let $p \in \mathcal{P}$. Let $\vec{\alpha}^{(1)}, \vec{\alpha}^{(2)} \in \mathbb{Z}^{\mathcal{P}}$. Then*

$$\begin{aligned}&\Big||\mathcal{B}_{p,V}(\vec{\alpha}^{(1)}, n)| - |\mathcal{B}_{p,V}(\vec{\alpha}^{(2)}, n)|\Big| \\ &\qquad\qquad \leq \left(\deg V \binom{n}{\dim V - 1} + O_V(n^{\dim V - 2})\right) \sum_{p' \in \mathcal{P}} \left|(\alpha_{p'}^{(1)} - \alpha_p^{(1)}) - (\alpha_{p'}^{(2)} - \alpha_p^{(2)})\right|.\end{aligned}$$

5.4. **Joints configuration.** We are ready to discuss joints of varieties. Here we set some notation and definitions.

By a $(k_1, \ldots, k_r; m_1, \ldots, m_r)$-*joints configuration* (or just a *joints configuration* for short) we mean a tuple $(\mathcal{J}, \mathcal{V}_1, \ldots, \mathcal{V}_d)$ as in Theorem 1.10, namely that each $\mathcal{V}_i$ is a finite multiset of k_i-dimensional varieties in $\mathbb{F}^d$, where $d = m_1k_1 + \cdots + m_rk_r$, and $\mathcal{J}$ is the set of joints formed by choosing m_i elements from $\mathcal{V}_i$ for each $i = 1, \ldots, r$. We write $\mathcal{M}(p)$ for the multiset of r-tuples $(\mathcal{S}_1, \ldots, \mathcal{S}_r)$, where each $\mathcal{S}_i$ is an unordered m_i-tuple of elements of $\mathcal{S}_i$ and such that together these $s = m_1 + \cdots + m_r$ varieties form a joint at p. The quantity $M(p)$ from Theorem 1.10 is then the cardinality of $\mathcal{M}(p)$. We have $M(p) > 0$ at each $p \in \mathcal{J}$.

5.5. **Vanishing lemma.** Before stating the analog to Lemma 3.9, let us first note the following observation about how high order directional derivatives of several varieties interact at a joint.

Lemma 5.7. *Let p be a joint formed by varieties $V_1, \ldots, V_s$. Suppose $g \in \mathbb{F}[x_1, \ldots, x_d]$ vanishes to order exactly r at p (as a polynomial function on $\mathbb{F}^d$). Then there exist $r_1, \ldots r_s \in \mathbb{Z}_{\geq 0}$ with $r_1 + \cdots + r_s = r$ and $D_1 \in \mathbb{D}_{p,V_1}^{r_1}, \ldots, D_s \in \mathbb{D}_{p,V_s}^{r_s}$ such that*

$$\left(D_1 D_2 \cdots D_s g\right)(p) \neq 0.$$

Proof. Let $k_i = \dim V_i$ for each i. By an affine change of coordinates, suppose that p is at the origin, V_1 is tangent to the first k_1 coordinate vectors, V_2 tangent to the next k_2 coordinate vectors, and so on. Let $cx_1^{\gamma_1} \cdots x_d^{\gamma_d}$, $c \in \mathbb{F} \setminus \{0\}$, be a monomial of lowest degree in g. Since g vanishes to order r at p, we have $\gamma_1 + \cdots + \gamma_d = r$. Let r_1 be the sum of the first k_1 γ_i's, r_2 the sum of the next k_2 γ_i's, and so on. By (4.2), there exist $D_i \in \mathbb{D}_{p,V_i}^{r_i}$ of the form

$$\begin{aligned}D_1 &= \mathsf{H}^{(\gamma_1, \ldots, \gamma_{k_1}, 0, \ldots, 0)} + \text{lower order derivatives}, \\ D_2 &= \mathsf{H}^{(0, \ldots, 0, \gamma_{k_1+1}, \ldots, \gamma_{k_1+k_2}, 0, \ldots, 0)} + \text{lower order derivatives}, \\ &\ldots.\end{aligned}$$

Then $D_1 D_2 \cdots D_s g = c +$ higher order terms, which evaluates to $c \neq 0$ at $p = 0$. ☐

The next statement is analogous to the vanishing lemma for planes in Lemma 3.9. The proof is analogous, but we write it out explicitly here since it is a critical step of the argument.

Lemma 5.8. *Let $(\mathcal{J}, \mathcal{V}_1, \ldots, \mathcal{V}_k)$ be a $(k_1, \ldots, k_r; m_1, \ldots, m_r)$-joints configuration. Let $s = m_1 + \cdots + m_r$ and $d = m_1k_1 + \cdots + m_rk_r$. Fix a handicap $\vec{\alpha}$ and its associated priority order. Fix a positive integer n. Choose $\mathcal{D}_{p,V}$ as earlier. For each $p \in \mathcal{J}$, fix a choice $V_1(p), V_2(p), \ldots, V_s(p)$ of varieties that form a joint at p, and of which exactly m_i of them come from $\mathcal{V}_i$ for each $i = 1, \ldots, r$.*

Then for every nonzero polynomial $g \in \mathbb{F}[x_1, \ldots, x_d]$ of degree at most n, one has $D_1 \cdots D_s g(p) \neq 0$ for some joint $p \in \mathcal{J}$ and some $D_1 \in \mathcal{D}_{p,V_1(p)}$, $\ldots$, $D_s \in \mathcal{D}_{p,V_s(p)}$.

Proof. Suppose, on the contrary, that there were some nonzero polynomial $g \in \mathbb{F}[x_1, \ldots, x_d]$ of degree at most n such that $D_1 \cdots D_s g(p) = 0$ for every joint $p \in \mathcal{J}$ and $D_1 \in \mathcal{D}_{p,V_1}$, $\ldots$, $D_s \in \mathcal{D}_{p,V_s}$, where $V_1, V_2, \ldots, V_s$ are any varieties that form a joint at p and exactly m_i of them come from $\mathcal{V}_i$ for each $i = 1, \ldots, r$, Choose $p \in \mathcal{J}$ to minimize $(p, v_p(g))$ under $\prec$, where $v_p(g)$ is the order vanishing of g at p.

Since g vanishes to order exactly $v_p(g)$ at p, by Lemma 5.7, there exist $D_1 \in \mathbb{D}^{r_1}_{p,V_1(p)}, \ldots, D_s \in \mathbb{D}^{r_s}_{p,V_s(p)}$ with $D_1 D_2 \cdots D_s g(p) \neq 0$ and $r_1 + \cdots + r_s = v_p(g)$. Among all choices of $D_1, \ldots, D_s$ (including choices of $r_1, \ldots, r_s$), choose ones so that $|\{i \in [s] : D_i \in \mathcal{D}_{p,V_i(p)}\}|$ is maximized. By the assumption at the beginning of the proof, one must have $D_i \notin \mathcal{D}_{p,V_i}$ for some $i \in [s]$. Relabeling if necessary, assume that $D_1 \notin \mathcal{D}_{p,V_1(p)}$. (Here we are using that derivatives commute.)

Suppose $p' \in V_1(p) \cap \mathcal{J}$ and $r' \in \mathbb{Z}_{\geq 0}$ satisfy $(p', r') \prec (p, r_1)$. We get $(p', r' + r_2 + \cdots + r_s) \prec (p, r_1 + r_2 + \cdots + r_s) = (p, v_p(g))$. By the choice of p, we have $(p, v_p(g)) \preceq (p', v_{p'}(g))$. Thus $(p', r' + r_2 + \cdots + r_s) \prec (p', v_{p'}(g))$, and hence $r' + r_2 + \cdots + r_s < v_{p'}(g)$. If follows that $DD_2 \cdots D_s g(p') = 0$ for all $D \in \mathbb{D}^{r'}_{p',V_1(p)}$ by the definition of vanishing order.

From the above paragraph we deduce that $D_2 \cdots D_s g(p')$ lies in the common kernel of $\mathbb{B}^{r'}_{p',V_1(p)}$ ranging over all $(p', r') \in (V_1(p) \cap \mathcal{J}) \times \mathbb{Z}_{\geq 0}$ with $(p', r') \prec (p, r_1)$. Since $D_1 D_2 \cdots D_s g(p) \neq 0$, we deduce that $D_2 \cdots D_s g$ does not lie in the common kernel of $\mathcal{B}^{r_1}_{p,V_1(p)}$, i.e., there is some $D \in \mathcal{D}^{r_1}_{p,V_1(p)}$ with $DD_2 \cdots D_s g(p) \neq 0$. But this D contradicts the earlier assumption that the choice of $D_1, \ldots, D_s$ maximizes $|\{i \in [s] : D_i \in \mathcal{D}_{p,V_i(p)}\}|$. $\square$

The next lemma is a consequence of parameter counting. Its proof is identical to that of Lemma 3.10 except that we now apply Lemma 5.8.

Lemma 5.9. *Assume the same setup as Lemma 5.8. We have*

$$\sum_{p \in \mathcal{J}} \prod_{i=1}^{s} \left|\mathcal{D}_{p,V_i(p)}(\vec{\alpha}, n)\right| \geq \binom{n+d}{d}.$$

5.6. **Choosing the handicaps.** We say that a joints configuration $(\mathcal{J}, \mathcal{V}_1, \ldots, \mathcal{V}_r)$ is *connected* if the following graph is connected: the vertex set is $\mathcal{J}$, with $p, p' \in \mathcal{J}$ adjacent if there is some $V \in \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_r$ containing both p and p'.

Lemma 5.10. *Let n be a positive integer and $(\mathcal{J}, \mathcal{V}_1, \ldots, \mathcal{V}_k)$ be a connected $(k_1, \ldots, k_r; m_1, \ldots, m_r)$-joints configuration. Let $\omega(p)$ be a positive real for each $p \in \mathcal{J}$. Then there exists a choice of handicap $\vec{\alpha} \in \mathbb{Z}^{\mathcal{J}}$ such that*

$$\frac{1}{\omega(p)} \left[\prod_{(\mathcal{S}_1, \ldots, \mathcal{S}_r) \in \mathcal{M}(p)} \prod_{V \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_r} \frac{|\mathcal{D}_{p,V}(\vec{\alpha}, n)|}{\binom{n}{\dim V}} \right]^{1/M(p)}$$

lies in some common interval of length $o_{\mathcal{J},\mathcal{V}_1,\ldots,\mathcal{V}_k,\omega;n\to\infty}(1)$ as we range over $p \in \mathcal{J}$. Here the notation means that the length of the interval tends to zero as n goes to infinity but the rate may depend on the joints configuration and ω.

Proof. The proof is analogous to Lemma 3.11 with appropriate modification. In this proof, we use $o(1)$ to denote $o_{\mathcal{J},\mathcal{V}_1,\ldots,\mathcal{V}_k,\omega;n\to\infty}(1)$. Let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence tending to 0 sufficiently slowly as n tends to infinity. Denote by $W_p(\vec{\alpha})$ the quantity (the dependence of $W_p(\vec{\alpha})$ on n is suppressed in the notation).

$$\frac{1}{\omega(p)} \left[\prod_{(\mathcal{S}_1, \ldots, \mathcal{S}_r) \in \mathcal{M}(p)} \prod_{V \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_r} \frac{|\mathcal{D}_{p,V}(\vec{\alpha}, n)|}{\binom{n}{\dim V}} \right]^{1/M(p)}.$$

We begin by noticing that, by Lemma 5.4, there exists some c depending on n and the joints configuration such that if $\alpha_p < \alpha_{p'} - c$ for two joints p, p' on the same flat V, then $|\mathcal{D}_{p,V}(\vec{\alpha}, n)| = 0$, which shows that $W_p(\vec{\alpha}) = 0$. Therefore, although there are infinitely many choices for $\vec{\alpha} \in \mathbb{Z}^{\mathcal{J}}$, there are only finitely many possible values of $(W_p)_{p\in\mathcal{J}}$ they can produce for a given n. Choose the $W_p(\vec{\alpha})$ so that after sorting W_p in the descending order, it has the least lexicographical order. Suppose that the sorted result is

$$W_{p_1}(\vec{\alpha}) \geq \cdots \geq W_{p_{|\mathcal{J}|}}(\vec{\alpha}).$$

It suffices to show that $W_{p_i}(\vec{\alpha}) - W_{p_{i+1}}(\vec{\alpha}) \leq \delta_n$ for all $i = 1, \dots, |\mathcal{J}| - 1$.

Suppose for the sake of contradiction that the claim fails for some i. Let t be the least positive integer such that $W_{p_t}(\vec{\alpha}) - W_{p_{t+1}}(\vec{\alpha}) > \delta_n$. Then let $\vec{v} = \vec{e}_{p_1} + \cdots + \vec{e}_{p_t}$ and $\vec{\alpha}' = \vec{\alpha} - \vec{v}$. Take a constant C' larger than all the degrees of the varieties in the joints configuration. Similar to the proof of Lemma 3.11, we can apply Lemma 5.6 to show that $||\mathcal{D}_{p,V}(\vec{\alpha}, n)| - |\mathcal{D}_{p,V}(\vec{\alpha}', n)|| / \binom{n}{\dim V} = o(1)$. Together with the fact that $|\mathcal{D}_{p,V}(\vec{\alpha}, n)| / \binom{n}{\dim V} \leq C' + o(1)$ (guaranteed by (5.1)) we can use a similar telescoping inequality to show that the difference between $W_p(\vec{\alpha}, n)$ and $W_p(\vec{\alpha}', n)$ is at most $o(1)$. Therefore the difference is bounded by $\delta_n/2$ as long as δ_n tends to 0 slowly enough.

Now, by the new monotonicity established in Lemma 5.5, we know that $W_{p_i}(\vec{\alpha}') \leq W_{p_i}(\vec{\alpha})$ if $i \leq t$, and $W_{p_i}(\vec{\alpha}') \geq W_{p_i}(\vec{\alpha})$ if $i > t$. If $W_p(\vec{\alpha}) \neq W_p(\vec{\alpha}')$ for some $p \in \mathcal{J}$, then by (5.1), we know that there exist $i \leq t$ and $p_i \in V$ such that $|\mathcal{D}_{p_i,V}(\vec{\alpha}', n)| < |\mathcal{D}_{p_i,V}(\vec{\alpha}, n)|$, resulting in $W_{p_i}(\vec{\alpha}') < W_{p_i}(\vec{\alpha})$. By the fact that $|W_p(\vec{\alpha}') - W_p(\vec{\alpha})| \leq \delta_n/2$ for all $p \in \mathcal{J}$ and the assumption that $W_{p_t} - W_{p_{t+1}} > \delta_n$, we know that $W_{p_1}(\vec{\alpha}'), \dots, W_{p_t}(\vec{\alpha}')$ are still the t largest ones among $(W_p(\vec{\alpha}'))_{p \in \mathcal{J}}$. Hence, that $W_{p_i}(\vec{\alpha}') < W_{p_i}(\vec{\alpha})$ is a contradiction with the assumption of the minimality under the lexicographical order.

The previous paragraph shows that $W_p(\vec{\alpha}) = W_p(\vec{\alpha}')$ for every $p \in \mathcal{J}$. As a consequence, $W_p(\vec{\alpha}) = W_p(\vec{\alpha} - m\vec{v})$ for all positive integers m and $p \in \mathcal{J}$. Since the joints configuration is connected, we can find $i \leq t < j$ such that p_i and p_j lie on the same variety. When m is sufficiently large, we have $\alpha_{p_i} - m < \alpha_{p_j} - c$, which forces $W_{p_i}(\vec{\alpha} - m\vec{v})$ to be 0. By the ordering, this shows that $W_{p'_i}(\vec{\alpha} - m\vec{v}) = 0$ for all $i' \leq i$. In particular, $W_{p_t}(\vec{\alpha}) = W_{p_{t+1}}(\vec{\alpha}) = 0$, contradicting $W_{p_t}(\vec{\alpha}) - W_{p_{t+1}}(\vec{\alpha}) > \delta_n$. □

Proof of Theorem 1.10(b). In this proof $o(1)$ denotes a quantity which goes zero as n goes to infinity but can dependent arbitrarily on the joints configuration. Similar to the proof of Theorem 1.2, it suffices to consider the case where the joints configuration is connected. Set

$$s = m_1 + \cdots + m_r,$$

and

$$J_\omega = \sum_{p \in \mathcal{J}} \omega(p) \quad \text{where} \quad \omega(p) = M(p)^{1/(s-1)}.$$

Choose $\vec{\alpha}$ according to Lemma 5.10. Then we can choose W so that

$$\left| \frac{1}{\omega(p)} \left[\prod_{(\mathcal{S}_1, \dots, \mathcal{S}_r) \in \mathcal{M}(p)} \prod_{V \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_r} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{\dim V}} \right]^{1/M(p)} - W \right| = o(1)$$

for all $p \in \mathcal{J}$. Hence, by Lemma 5.9 (choosing $V_i(p)$ of Lemma 5.9 to give the minimum product below), we have that

$$\sum_{p \in \mathcal{J}} \omega(p) W \geq \sum_{p \in \mathcal{J}} \min_{(\mathcal{S}_1, \dots, \mathcal{S}_r) \in \mathcal{M}(p)} \prod_{V \in \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_r} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{\dim V}} - o(1) \geq \frac{\binom{n+d}{d}}{\prod_{i=1}^r \binom{n}{k_i}^{m_i}} - o(1),$$

which, after rearrangement, shows that

$$W \geq \frac{\prod_{i=1}^r (k_i!)^{m_i}}{J_\omega \cdot d!} - o(1).$$

Let $\mathcal{V}_{p,i}$ be the set of varieties in $\mathcal{V}_i$ that contain p. Then we have that for any joint $p \in \mathcal{J}$,

$$
\begin{aligned}
M(p)\omega(p)W &\le \sum_{(\mathcal{S}_1,\dots,\mathcal{S}_r)\in\mathcal{M}(p)} \prod_{V\in\mathcal{S}_1\cup\cdots\cup\mathcal{S}_r} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{\dim V}} + o(1) && \text{[by AM-GM]}\\
&\le \sum_{\mathcal{S}_1\in\binom{\mathcal{V}_{p,1}}{m_1},\dots,\mathcal{S}_r\in\binom{\mathcal{V}_{p,r}}{m_r}} \prod_{i=1}^r \prod_{V\in\mathcal{S}_i} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{k_i}} + o(1)\\
&= \prod_{i=1}^r \sum_{\mathcal{S}_i\in\binom{\mathcal{V}_{p,i}}{m_i}} \prod_{V\in\mathcal{S}_i} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{k_i}} + o(1)\\
&\le \prod_{i=1}^r \frac{(\deg \mathcal{V}_i)^{m_i}}{m_i!} \left(\sum_{V\in\mathcal{V}_{p,i}} \frac{|\mathcal{D}_{p,V}|}{\deg \mathcal{V}_i \binom{n}{k_i}} \right)^{m_i} + o(1)\\
&\le \frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r! s^s} \left(\sum_{i=1}^r \sum_{V\in\mathcal{V}_{p,i}} \frac{m_i\,|\mathcal{D}_{p,V}|}{\deg \mathcal{V}_i \binom{n}{k_i}} \right)^{s} + o(1). && \text{[by AM-GM]}
\end{aligned}
$$

By taking the s-th root on both sides, summing over all p using (5.1) and noticing that $M(p)\omega(p) = \omega(p)^s$, we conclude that

$$
\begin{aligned}
\sum_{p\in\mathcal{J}} \omega(p) W^{1/s} &\le \frac{1}{s} \left(\frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r!} \right)^{1/s} \sum_{i=1}^r \sum_{V\in\mathcal{V}_i} \sum_{p\in\mathcal{J}\cap\mathcal{V}_i} \frac{m_i\,|\mathcal{D}_{p,V}|}{\deg \mathcal{V}_i \binom{n}{k_i}} + o(1)\\
&= \frac{1}{s} \left(\frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r!} \right)^{1/s} \sum_{i=1}^r \sum_{V\in\mathcal{V}_i} \frac{m_i \deg V}{\deg \mathcal{V}_i} + o(1)\\
&= \frac{1}{s} \left(\frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r!} \right)^{1/s} \sum_{i=1}^r m_i + o(1)\\
&= \left(\frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r!} \right)^{1/s} + o(1).
\end{aligned}
$$

Rearranging, we find

$$
W \le \frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r! J_\omega^s} + o(1).
$$

By comparing the lower and upper bounds on W, and letting $n \to \infty$ so that the $o(1)$ term vanishes, we have

$$
\frac{\prod_{i=1}^r (k_i!)^{m_i}}{J_\omega \cdot d!} \le \frac{(\deg \mathcal{V}_1)^{m_1}\cdots(\deg \mathcal{V}_r)^{m_r}}{m_1!\cdots m_r! J_\omega^s}.
$$

Rearranging gives the desired conclusion

$$
J_\omega \le \left(d! \prod_{i=1}^r \frac{(\deg \mathcal{V}_i)^{m_i}}{k_i!^{m_i} m_i!} \right)^{1/(s-1)}. \qquad \square
$$

Proof of Theorem 1.10(a). As earlier, we may assume that the joints configuration is connected. Set $s = m_1 + \cdots + m_r$ throughout the proof. Choose $\vec{\alpha}$ according to Lemma 5.10 with $\omega(p) = 1$ for all $p \in \mathcal{J}$. Then we can choose W so that

$$
\left| \left[\prod_{(\mathcal{S}_1,\dots,\mathcal{S}_r)\in\mathcal{M}(p)} \prod_{V\in\mathcal{S}_1\cup\cdots\cup\mathcal{S}_r} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{\dim V}} \right]^{1/M(p)} - W \right| = o(1)
$$

for all $p \in \mathcal{J}$. Hence, by Lemma 5.9, we have that

$$\sum_{p\in\mathcal{J}} W \geq \sum_{p\in\mathcal{J}} \min_{(\mathcal{S}_1,\dots,\mathcal{S}_r)\in\mathcal{M}(p)} \prod_{V\in\mathcal{S}_1\cup\cdots\cup\mathcal{S}_r} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{\dim V}} - o(1) \geq \frac{\binom{n+d}{d}}{\prod_{i=1}^r \binom{n}{k_i}^{m_i}} - o(1),$$

which, after rearrangement, shows that

$$W \geq \frac{\prod_{i=1}^r (k_i!)^{m_i}}{|\mathcal{J}|\cdot d!} - o(1).$$

For each $p \in \mathcal{J}$, let $(\mathcal{S}_1(p),\dots,\mathcal{S}_r(p)) \in \mathcal{M}(p)$ be the element S of $\mathcal{M}(p)$ such that

$$\prod_{V\in\mathcal{S}_1\cup\cdots\cup\mathcal{S}_r} \frac{|\mathcal{D}_{p,V}|}{\binom{n}{\dim V}}$$

is maximized. Then $W \leq \prod_{V\in\mathcal{S}_1(p)\cup\cdots\cup\mathcal{S}_r(p)} |\mathcal{D}_{p,V}|/\binom{n}{\dim V} + o(1)$, which shows that

$$\begin{aligned}
&s\,|\mathcal{J}|\,W^{1/s}\,(\deg\mathcal{V}_1)^{-m_1/s}\cdots(\deg\mathcal{V}_r)^{-m_r/s} \\
&\qquad \leq s\sum_{p\in\mathcal{J}}\left(\prod_{i=1}^r m_i^{-m_i}\prod_{V\in\mathcal{S}_i(p)}\frac{m_i\,|\mathcal{D}_{p,V}|}{\deg\mathcal{V}_i\binom{n}{k_i}}\right)^{1/s} + o(1) \\
&\qquad \leq \frac{1}{m_1^{m_1/s}\cdots m_r^{m_r/s}}\sum_{p\in\mathcal{J}}\sum_{i=1}^r\sum_{V\in\mathcal{S}_i(p)}\frac{m_i\,|\mathcal{D}_{p,V}|}{\deg\mathcal{V}_i\binom{n}{k_i}} + o(1) && \text{[by AM-GM]} \\
&\qquad \leq \frac{1}{m_1^{m_1/s}\cdots m_r^{m_r/s}}\sum_{i=1}^r\sum_{V\in\mathcal{V}_i}\sum_{p\in\mathcal{J}\cap V}\frac{m_i\,|\mathcal{D}_{p,V}|}{\deg\mathcal{V}_i\binom{n}{k_i}} + o(1) \\
&\qquad \leq \frac{1}{m_1^{m_1/s}\cdots m_r^{m_r/s}}\sum_{i=1}^r\sum_{V\in\mathcal{V}_i}\frac{m_i\deg V}{\deg\mathcal{V}_i} + o(1) && \text{[by (5.1)]} \\
&\qquad \leq \frac{1}{m_1^{m_1/s}\cdots m_r^{m_r/s}}\sum_{i=1}^r m_i + o(1) \\
&\qquad = \frac{s}{m_1^{m_1/s}\cdots m_r^{m_r/s}} + o(1).
\end{aligned}$$

By rearranging, we get that

$$W \leq \frac{(\deg\mathcal{V}_1)^{m_1}\cdots(\deg\mathcal{V}_r)^{m_r}}{m_1^{m_1}\cdots m_r^{m_r}\,|\mathcal{J}|^s} + o(1).$$

By comparing the lower and upper bounds on W, and letting $n\to\infty$ so that the $o(1)$ term vanishes, we have

$$\frac{\prod_{i=1}^r (k_i!)^{m_i}}{|\mathcal{J}|\cdot d!} \leq \frac{(\deg\mathcal{V}_1)^{m_1}\cdots(\deg\mathcal{V}_r)^{m_r}}{m_1^{m_1}\cdots m_r^{m_r}\,|\mathcal{J}|^s}.$$

Rearranging gives the desired conclusion

$$|\mathcal{J}| \leq \left(d!\prod_{i=1}^r \frac{(\deg V_i)^{m_i}}{k_i!^{m_i} m_i^{m_i}}\right)^{1/(s-1)}. \qquad \square$$

Acknowledgments

This research was conducted while Yu was a participant and Tidor was a mentor in the Summer Program in Undergraduate Research (SPUR) of the MIT Mathematics Department. We thank David Jerison and Ankur Moitra for their role in advising the program. We also thank the anonymous referee for a careful reading of the manuscript and helpful comments.

Tidor was supported by the NSF Graduate Research Fellowship Program DGE-1745302.

Zhao was supported by NSF Award DMS-1764176, the MIT Solomon Buchsbaum Fund, and a Sloan Research Fellowship.

References

[1] Jonathan Bennett, Anthony Carbery, and Terence Tao, *On the multilinear restriction and Kakeya conjectures*, Acta Math. **196** (2006), 261–302.

[2] Pavle V. M. Blagojević, Aleksandra S. Dimitrijević Blagojević, and Günter M. Ziegler, *Polynomial partitioning for several sets of varieties*, J. Fixed Point Theory Appl. **19** (2017), 1653–1660.

[3] Boris Bukh and Ting-Wei Chao, *Sharp density bounds on the finite field Kakeya problem*, Discrete Anal. (2021), Paper No. 26, 9pp.

[4] Anthony Carbery and Marina Iliopoulou, *Counting joints in vector spaces over arbitrary fields*, arXiv:1403.6438.

[5] Anthony Carbery and Marina Iliopoulou, *Joints formed by lines and a k-plane, and a discrete estimate of Kakeya type*, Discrete Anal. (2020), Paper No. 18, 45pp.

[6] Anthony Carbery and Stefán Ingi Valdimarsson, *Colouring multijoints*, Discrete Comput. Geom. **52** (2014), 730–742.

[7] Bernard Chazelle, Herbert Edelsbrunner, Leonidas J. Guibas, Richard Pollack, Raimund Seidel, Micha Sharir, and Jack Snoeyink, *Counting and cutting cycles of lines and rods in space*, Comput. Geom. **1** (1992), 305–323.

[8] Zeev Dvir, *On the size of Kakeya sets in finite fields*, J. Amer. Math. Soc. **22** (2009), 1093–1097.

[9] Zeev Dvir, *Incidence theorems and their applications*, Found. Trends Theor. Comput. Sci. **6** (2010), 257–393 (2012).

[10] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan, *Extensions to the method of multiplicities, with applications to Kakeya sets and mergers*, SIAM J. Comput. **42** (2013), 2305–2328.

[11] David Eisenbud, *Commutative algebra. with a view toward algebraic geometry*, Graduate Texts in Mathematics, vol. 150, Springer-Verlag, New York, 1995.

[12] György Elekes, Haim Kaplan, and Micha Sharir, *On lines, joints, and incidences in three dimensions*, J. Combin. Theory Ser. A **118** (2011), 962–977.

[13] Larry Guth, *The endpoint case of the Bennett-Carbery-Tao multilinear Kakeya conjecture*, Acta Math. **205** (2010), 263–286.

[14] Larry Guth, *Polynomial partitioning for a set of varieties*, Math. Proc. Cambridge Philos. Soc. **159** (2015), 459–469.

[15] Larry Guth, *Polynomial methods in combinatorics*, University Lecture Series, vol. 64, American Mathematical Society, Providence, RI, 2016.

[16] Larry Guth, *A restriction estimate using polynomial partitioning*, J. Amer. Math. Soc. **29** (2016), 371–413.

[17] Larry Guth, *Restriction estimates using polynomial partitioning II*, Acta Math. **221** (2018), 81–142.

[18] Larry Guth and Nets Hawk Katz, *Algebraic methods in discrete analogs of the Kakeya problem*, Adv. Math. **225** (2010), 2828–2839.

[19] Larry Guth and Nets Hawk Katz, *On the Erdős distinct distances problem in the plane*, Ann. of Math. (2) **181** (2015), 155–190.

[20] Márton Hablicsek, *On the joints problem with multiplicities*, arXiv:1408.5791.

[21] Marina Iliopoulou, *Counting joints with multiplicities*, Proc. Lond. Math. Soc. (3) **106** (2013), 675–702.

[22] Marina Iliopoulou, *Discrete analogues of kakeya problems*, 2013, Ph.D. Thesis, University of Edinburgh, arXiv:1312.5436.

[23] Marina Iliopoulou, *Counting multijoints*, J. Combin. Theory Ser. A **136** (2015), 143–163.

[24] Marina Iliopoulou, *Incidence bounds on multijoints and generic joints*, Discrete Comput. Geom. **54** (2015), 481–512.

[25] Haim Kaplan, Micha Sharir, and Eugenii Shustin, *On lines and joints*, Discrete Comput. Geom. **44** (2010), 838–843.

[26] René Quilodrán, *The joints problem in* $\mathbb{R}^n$, SIAM J. Discrete Math. **23** (2009/10), 2211–2213.

[27] József Solymosi and Terence Tao, *An incidence theorem in higher dimensions*, Discrete Comput. Geom. **48** (2012), 255–280.

[28] Endre Szemerédi and William T. Trotter, Jr., *Extremal problems in discrete geometry*, Combinatorica **3** (1983), 381–392.

[29] Terence Tao, *Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory*, EMS Surv. Math. Sci. **1** (2014), 1–46.

[30] Terry Tao, *A partial converse to Bezout's theorem*, blog post `https://terrytao.wordpress.com/2012/09/25/a-partial-converse-to-bezouts-theorem/`.

[31] Ravi Vakil, *The rising sea: Foundations of algebraic geometry*, `http://math.stanford.edu/~vakil/216blog/FOAGnov1817public.pdf`, November 18, 2017 draft.

[32] Miguel N. Walsh, *Concentration estimates for algebraic intersections*, arXiv:1906.05843.

[33] Miguel N. Walsh, *The polynomial method over varieties*, Invent. Math. **222** (2020), 469–512.

[34] Thomas Wolff, *Recent work connected with the Kakeya problem*, Prospects in mathematics (Princeton, NJ, 1996), Amer. Math. Soc., Providence, RI, 1999, pp. 129–162.

[35] Ben Yang, *Generalizations of joints problem*, arXiv:1606.08525.

[36] Ben Yang, *Polynomial Partitioning and Incidence Problems in Higher Dimensions*, 2017, Ph.D. thesis, Massachusetts Institute of Technology, `https://dspace.mit.edu/handle/1721.1/112880`.

[37] Hung-Hsun Hans Yu and Yufei Zhao, *Joints tightened*, Amer. J. Math., to appear, arXiv:1911.08605.

[38] Ruixiang Zhang, *The endpoint perturbed Brascamp-Lieb inequalities with examples*, Anal. PDE **11** (2018), 555–581.

[39] Ruixiang Zhang, *A proof of the Multijoints Conjecture and Carbery's generalization*, J. Eur. Math. Soc. (JEMS) **22** (2020), 2405–2417.

Tidor, Zhao: Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Yu: Trinity College, University of Cambridge, UK
Email address: `{jtidor,hansonyu,yufeiz}@mit.edu`