

On Sets Defining Few Ordinary Lines

Ben Green · Terence Tao

Received: 24 August 2012 / Revised: 14 May 2013 / Accepted: 21 May 2013 /
Published online: 27 June 2013
© Springer Science+Business Media New York 2013

Abstract Let P be a set of n points in the plane, not all on a line. We show that if n is large then there are at least $n/2$ *ordinary* lines, that is to say lines passing through exactly two points of P . This confirms, for large n , a conjecture of Dirac and Motzkin. In fact we describe the exact extremisers for this problem, as well as all sets having fewer than $n - C$ ordinary lines for some absolute constant C . We also solve, for large n , the “orchard-planting problem”, which asks for the maximum number of lines through exactly 3 points of P . Underlying these results is a structure theorem which states that if P has at most Kn ordinary lines then all but $O(K)$ points of P lie on a cubic curve, if n is sufficiently large depending on K .

Keywords Sylvester–Gallai · Ordinary lines · Cubic curves · Dirac–Motzkin

Mathematics Subject Classification (1991) 20G40 · 20N99

1 Introduction

The *Sylvester–Gallai theorem* is a well-known theorem in combinatorial geometry. It was proven by Gallai [18] in response to a question of Sylvester [33] from 40 years earlier (Fig. 1).

B. Green (✉)
Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA,
England, UK
e-mail: b.j.green@dpmms.cam.ac.uk

T. Tao
Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles, CA 90095,
USA
e-mail: tao@math.ucla.edu

QUESTIONS FOR SOLUTION.

11851. (Professor SYLVESTER.)—Prove that it is not possible to arrange any finite number of real points so that a right line through every two of them shall pass through a third, unless they all lie in the same right line.

Fig. 1 Sylvester's question [33]

Theorem 1.1 (Sylvester–Gallai theorem) *Suppose that P is a finite set of points in the plane, not all on one line. Then there is an ordinary line spanned by P , that is to say a line in P containing exactly two points.*

Several different proofs of this now appear in the literature. We will be particularly interested in a proof due to Melchior [25] based on projective duality and Euler's formula, which we will recall in Sect. 3. It is natural to wonder how many ordinary lines there are in a set P of points, not all on a line, when the cardinality $|P|$ of P is equal to n . Melchior's argument in fact shows that there are at least three ordinary lines, but considerably more is known. Motzkin [26] was the first to obtain a lower bound (of order $n^{1/2}$) tending to infinity with n . Kelly and Moser [22] proved that there are at least $3n/7$ ordinary lines, and Csima and Sawyer [11] improved this to $6n/13$ when $n > 7$. Their work used ideas from the thesis of Hansen [20], which purported to prove the $n/2$ lower bound but was apparently flawed. An illuminating discussion of this point may be found in the MathSciNet review of [11]. There are several nice surveys on this and related problems; see [4], [15, Chap. 17], [27] or [28].

One of our main objectives in this paper is to clarify this issue for large n . The following theorem resolves, for large n , a long-standing conjecture which has been known as the *Dirac–Motzkin conjecture*. Apparently neither author formally conjectures this in print, though Dirac [12] twice states that its truth is “likely”. Motzkin [26] does not seem to mention it at all.

Theorem 1.2 (Dirac–Motzkin conjecture) *Suppose that P is a finite set of n points in the plane, not all on one line. Suppose that $n \geq n_0$ for a sufficiently large absolute constant n_0 . Then P spans at least $n/2$ ordinary lines.*

We will in fact establish a more precise result obtaining the exact minimum for all $n \geq n_0$ as well as a classification of the extremal examples. One rather curious feature of this more precise result is that if n is odd there must be at least $3\lfloor n/4 \rfloor$ ordinary lines. See Theorems 2.2 and 2.4 below for more details. When n is even, one can attain $n/2$ ordinary lines by combining $n/2$ equally spaced points on a circle with $n/2$ points at infinity; see Proposition 2.1 below.

For small values of n , there are exceptional configurations with fewer than $n/2$ ordinary lines. Kelly and Moser [22] observe that a triangle together with the midpoints of its sides and its centroid has $n = 7$ and just 3 ordinary lines. Crowe and McKee [10] provide a more complicated configuration with $n = 13$ and 6 ordinary lines. It is possible that Theorem 1.2 remains true for *all* n with the exception of these two examples (or equivalently, one could take n_0 as low as 14). Unfortunately our method does not give a good bound for n_0 ; we could in principle compute such a bound, but it would be of double exponential type and, we think, not worth the effort.

Fig. 2 Jackson’s question [21]

**2. Fain would I plant a grove in rows,
 But how must I its form compose
 With three trees in each row ;
 To have as many rows as trees ;
 Now tell me, artists, if you please ;
 ’Tis all I want to know.**

Our methods also apply (in fact with considerably less effort) to resolve sufficiently large instances of a slightly less well-known (but considerably older) problem referred to in the literature as the *orchard problem*. This was first formally posed by Sylvester [32] in 1868 (Fig. 3), though the 1821 book of Jackson [21] has a whole section containing puzzles of a similar flavour, posed more eloquently than any result in this paper (Fig. 2).

Theorem 1.3 (Orchard problem) *Suppose that P is a finite set of n points in the plane. Suppose that $n \geq n_0$ for some sufficiently large absolute constant n_0 . Then there are no more than $\lfloor n(n - 3)/6 \rfloor + 1$ lines that are 3-rich, that is they contain precisely 3 points of P . (Here and in the sequel, $\lfloor x \rfloor$ denotes the integer part of x .)*

This theorem is tight for large n , as noted by Sylvester [32], and also subsequently by Burr et al. [5], who discuss this problem extensively. We will give these examples, which are based on irreducible cubic curves, in Proposition 2.6 below. In fact these are the *only* examples where equality occurs for large n : see the remarks at the very end of Sect. 9. Again, there are counterexamples for small n . In particular, the example of a triangle, the midpoints of its sides and its centroid has $n = 7$ but 6 lines containing precisely three points of P ; by contrast, the bound of Theorem 1.3 is 5 in this case.

As observed in [5], lower bounds for the number N_2 of ordinary lines can be converted into upper bounds for the number N_3 of 3-rich lines thanks to the obvious double-counting identity $\sum_{k=2}^n \binom{k}{2} N_k = \binom{n}{2}$ (with N_k denoting the number of k -rich lines). In particular, previously known lower bounds on the Dirac–Motzkin conjecture can be used to deduce upper bounds on the orchard problem. However, one cannot deduce Theorem 1.3 in this fashion from Theorem 1.2; this is related to the fact that the extremal examples showing the sharpness of the two theorems are quite different, as we shall see in Sect. 2 below.

Underlying the proof of both of these results are structure theorems for sets with few ordinary lines, which are perhaps of independent interest. The most basic such result is the following. We use the asymptotic notation $X = O(Y)$ or $X \ll Y$ to denote the bound $|X| \leq CY$ for some absolute constant C .

Theorem 1.4 (Weak structure theorem) *Suppose that P is a finite set of n points in the plane. Suppose that P spans at most Kn ordinary lines for some $K \geq 1$. Suppose also that $n \geq \exp \exp(CK^C)$, for some sufficiently large absolute constant C . Then all but at most $O(K^{O(1)})$ points of P lie on an algebraic curve γ of degree at most 3.*

In fact we establish a slightly more precise statement, see Proposition 6.14 below. Note that we do not require the algebraic curve γ to be irreducible; thus γ could be an irreducible cubic, the union of a conic and a line, or the union of three lines. As

QUESTIONS FOR SOLUTION.

2572. (Proposed by Professor SYLVESTER.)—1. If h, k, l, \dots are the distinct prime factors of n , prove that the number of points (which I term pluperfect points of the n th order) at which a cubic curve can have the highest degree of contact with a curve of the degree n (not composed of repetitions of a curve of lower degree) is $9n^2(1-h^{-2})(1-k^{-2})(1-l^{-2})$.

2. Show that if the first tree in the solution to Quest. 2473 (*Reprint*, Vol. VIII., p. 106) be planted at a pluperfect point of the n th order in a cubic, the sequence of tree-marks 1, 2, 4, 5, 7... may be replaced by recurring periods of $2n$ numerals, and that the two halves of each period will consist of the same n numerals arranged in reverse order, that in fact only the first n of the numerals 1, 2, 4, ... need appear in the result.

3. Hence prove that n trees may be so arranged as to contain between them $E\left\{\frac{1}{3}n(n-1)\right\} - E\left\{\frac{1}{3}n\right\}$ rows of three in a row, where E (the symbol of entirety) denotes that the integer part only is to be taken of the function which it governs.

NOTE.—Thus, for 81 trees the number will be 1053 instead of 800, the number obtained when the first tree is at a non-pluperfect point; and for 15 trees the number is 30 instead of 26, the number stated in the Editorial reference at the end of the solution above cited as applicable to 15 points, but which is, in fact, the number for 14, according to the formula stated above. It must, however, be observed, that this formula, does not in general give the absolute maximum. Thus, for $n=9$ the formula gives only 9 rows, whereas the true maximum (dealing with *real* trees) is 10. In fact, $E\left\{\frac{1}{3}n(n-1)\right\}$ is the arithmetical limit obtained by dividing the number of duads of n by 3, the number of them in each row. Thus, $E\left\{\frac{1}{3}n\right\}$ is the difference between this limit and the actual number of the formula, which may also be expressed as equal to $\frac{1}{3}(n^2-3n)$ when this is an integer, and the integer immediately superior to it when it is fractional. Thus we see in a moment that with 1000 trees, the number of rows containing three in a row may be made equal to 166167. In a word, when the first tree is planted at an ordinary point in the cubic, the number of rows given by the formulæ in Quest. 2473 is $E\left\{\frac{1}{3}(n-1)^2\right\}$, when at a pluperfect point the number is $E\left\{\frac{1}{3}(n-1)(n-2)\right\}$.

Another method of placing points on a cubic which gives 10 rows for 9 points, and possibly in all cases the true maximum, will form the subject of a future question.

Fig. 3 Sylvester's question [32]

we shall see in later sections, cubic curves arise naturally in the study of point-line configurations with few ordinary lines, in large part due to the well-known abelian group structure (or pseudo-group structure) defined by the collinearity relation on such curves (or equivalently, by Chasles's version of the Cayley–Bacharach theorem, see Proposition 4.1). The lower bound $n \geq \exp \exp(CK^C)$ is present for rather artificial reasons, and can likely be improved substantially.

Projective geometry Much of the paper is best phrased in the language of projective geometry. We recall for the convenience of the reader the notion of the projective plane \mathbb{RP}^2 as $(\mathbb{R}^3 \setminus \{0\}) / \sim$, where $(x, y, z) \sim (x', y', z')$ if and only if there is some $\lambda \neq 0$ such that $x' = \lambda x, y' = \lambda y$ and $z' = \lambda z$. We denote points of \mathbb{RP}^2 with square brackets, thus $[x, y, z]$ is the equivalence class of (x, y, z) under \sim . We have the embedding $\mathbb{R}^2 \hookrightarrow \mathbb{RP}^2$ given by $(x, y) \mapsto [x, y, 1]$; in fact \mathbb{RP}^2 may be thought of as \mathbb{R}^2 together with the *line at infinity* consisting of points $[x, y, 0]$ (modulo the equivalence relation \sim). For the point-line incidence problems considered in this

paper, the projective and affine formulations are equivalent. Indeed given a finite set of points P in \mathbb{RP}^2 , we may apply a generic projective transformation so as to move all points of P to the affine plane \mathbb{R}^2 if desired, without affecting the number of ordinary lines or 3-rich lines. This is illustrated in the figures in Sect. 2.

For our purposes, there are two main advantages of working in projective space instead of affine space. The first is to allow the use of projective transformations to normalise one’s geometric configurations, for instance by moving a line to the line at infinity, or transforming a non-singular irreducible cubic curve into an elliptic curve in Weierstrass normal form. The other main advantage is the ability to utilise *point-line duality*. Given a point $p = [a, b, c]$, one may associate the line $p^* := \{[x, y, z] : ax + by + cz = 0\}$, and conversely given a projective line $\ell = \{[x, y, z] : ax + by + cz = 0\}$ one may associate the point $\ell^* = [a, b, c]$. It is clear that $p \in \ell$ if and only if $\ell^* \in p^*$. Working in the dual can provide us with information that is difficult to access otherwise. We shall see this twice: once in Sects. 3 and 4, when we apply Euler’s formula in the dual setting following an argument of Melchior [25], and then again in Sect. 6 where we will employ a convexity argument, due to Luke Alexander Betts, in the dual setting.

Next, we give a structure theorem which is more precise than Theorem 1.4.

Theorem 1.5 (Full structure theorem) *Suppose that P is a finite set of n points in the projective plane \mathbb{RP}^2 . Let $K > 0$ be a real parameter. Suppose that P spans at most Kn ordinary lines. Suppose also that $n \geq \exp \exp(CK^C)$ for some sufficiently large absolute constant C . Then, after applying a projective transformation if necessary, P differs by at most $O(K)$ points (which can be added or deleted) from an example of one of the following three types:*

- (i) $n - O(K)$ points on a line;
- (ii) The set

$$X_{2m} := \left\{ \left[\cos \frac{2\pi j}{m}, \sin \frac{2\pi j}{m}, 1 \right] : 0 \leq j < m \right\} \cup \left\{ \left[-\sin \frac{\pi j}{m}, \cos \frac{\pi j}{m}, 0 \right], 0 \leq j < m \right\} \tag{1.1}$$

consisting of m points on the unit circle and m points on the line at infinity, for some $m = \frac{n}{2} + O(K)$;

- (iii) A coset $H \oplus g, 3g \in H$, of a finite subgroup H of the non-singular real points on an irreducible cubic curve, with H having cardinality $n + O(K)$ (the group law \oplus on such curves is reviewed in Sect. 2 below).

Conversely, every set of this type has at most $O(Kn)$ ordinary lines.

We have the following consequence, which can handle slowly growing values of K .

Corollary 1.6 *Suppose that P is a finite set of n points in the projective plane \mathbb{RP}^2 . Suppose that P spans at most $n(\log \log n)^c$ ordinary lines for some sufficiently small constant $c > 0$. Then, after applying a projective transformation, P differs by at most $o(n)$ points from one of the examples (i), (ii), (iii) detailed in Theorem 1.5 above. In*

particular, we may add/remove $o(n)$ points to/from P to get a set with at most $n + O(1)$ ordinary lines.

Here, of course, $o(n)$ denotes a quantity which, after dividing by n , tends to zero as n goes to infinity.

Remark This corollary may, for all we know, be true with a much weaker assumption, perhaps even that P spans $o(n^2)$ ordinary lines. Very likely, if one could weaken the hypothesis $n \geq \exp \exp(CK^C)$ in Theorem 1.4 then one could do so also in Theorem 1.5.

Proof methods As mentioned previously, the starting point¹ of our arguments will be Melchior’s proof [25] of the Sylvester–Gallai theorem using duality and the Euler formula $V - E + F = 1$ for polygonal decompositions of the projective plane \mathbb{RP}^2 . Melchior’s argument uses at one point the obvious fact that all polygons have at least three sides to obtain an inequality implying the existence of ordinary lines. The same argument also shows that if a point set P spans very few ordinary lines, then almost all of the polygons in the dual configuration Γ_P (cut out by the dual lines p^* for $p \in P$) must in fact have *exactly* three sides. Because of this, it is possible to show in this case that the dual configuration contains large regions which have the combinatorial structure of a regular triangular grid.

The next key observation is that inside any triangular grid of non-trivial size, one can find “hexagonal” configurations of lines and points (see Fig. 14) which are dual to the configuration of lines and points arising in Chasles’s version of the Cayley–Bacharach theorem (Proposition 4.1 below). From this observation and some elementary combinatorial arguments, one can start placing large subsets of P on a single cubic curve. For instance, in Proposition 5.1 we will be able to establish a “cheap structure theorem” asserting that a set of n points with fewer than Kn ordinary lines can be covered by no more than $500K$ cubic curves. This observation turns out not to be new—a closely related technique is used in a paper of Carnicer and Godés [7] concerning *generalised principal lattices*, which arise in interpolation theory.

In principle, this cheap structure theorem already reduces the underlying geometry from a two-dimensional one (the projective plane \mathbb{RP}^2) to a one-dimensional one (the union of a number of cubic curves). Unfortunately, the collinearity relation between distinct cubic curves is too complicated to handle directly. Because of this, we must refine the previous combinatorial analysis to strengthen the structural control on a point set P with few ordinary lines. By studying the lines connecting a typical point p in P with all the other points in P using the triangular grid analysis, one can obtain a more complicated partition of P into cubic curves passing through p . A detailed statement may be found in Lemma 5.2. Comparing such a partition with the reference partition coming from the cheap structure theorem, one can obtain Proposition 5.3, a structure theorem of intermediate strength. Roughly speaking this result asserts that most of the points in P lie on a *single* irreducible cubic curve, on a union of an irreducible

¹ One defect of this approach is that it breaks down totally in the complex case, and so we have nothing new to say here about ordinary lines or 3-rich lines for configurations of complex lines in \mathbb{CP}^2 .

conic and a bounded number of lines (with the points shared almost evenly between the conic and the lines), or on the union of a bounded number of lines only.

The next stage is to cut the number of lines involved down to one. The key proposition² here is Proposition 6.3. It asserts, roughly speaking, that a set of n points on two or more lines, which contains $\gg n$ points on each line, must generate $\gg n^2$ ordinary lines. This is a statement that fails in finite field geometries, and must use at some point the torsion-free nature of the real line \mathbb{R} [see Sect. 2 for more discussion of this point, particularly with regards to the “near-counterexamples” (2.1) and (2.2)]. There are two key cases of the proposition which need to be established. The first is when the lines involved are all concurrent or, after a projective transformation, all parallel. In this case we use an argument of Betts, Proposition 6.4, involving projective duality and convexity. The use of convexity here is where the torsion-free nature of \mathbb{R} is implicitly used. In the case when the lines are not concurrent, we instead rely on Menelaus’s theorem to introduce various ratios of lengths, and then exploit a sum-product estimate of Elekes et al. [14]. This latter result, stated in Proposition A.9, also implicitly exploits the torsion-free nature of \mathbb{R} .

The result of the above analysis is a yet stronger structure theorem for sets P with few ordinary lines: P is mostly placed in either an irreducible cubic curve, the union of an irreducible conic and a line, or on a single line. A detailed statement may be found in Proposition 6.14. The latter case, in which almost all points lie on a line, is easily studied. To deal with the other two cases one uses the abelian group structure on irreducible cubics, as well as the analogous pseudo-group structure on the union of a conic and a line. The information that P contains few ordinary lines can then be converted to an additive combinatorics property on finite subsets of an abelian group. Fairly standard tools from additive combinatorics then show that P is almost a finite subgroup of that abelian group. This allows us to rule out “essentially torsion-free” situations, such as that provided by singular cubic curves (except for the acnodal singular cubic curve), and eventually leads to the full structure theorem in Theorem 1.5.

To solve the Dirac–Motzkin conjecture and the orchard problem for large n , we observe that potential counterexamples P to either conjecture will have few ordinary lines and hence can be described by Theorem 1.5. This quickly implies that P is close, up to projective transformation, to one of the known extremisers coming from roots of unity or from subgroups of elliptic curves, with a small number of additional points added or removed. The remaining task is to compute the effect that these added/removed points have on the number of ordinary lines or 3-rich lines in P . Here, to get the strongest results, we will need a slight variant of a result of Poonen and Rubinstein [29] in order to control the number of times a point may be concurrent with two roots of unity. See Proposition 7.5 for details.

Of the two problems, the orchard problem turns out to be somewhat easier, and can be in fact established using only the intermediate structure theorem in Proposition 5.3 rather than the more difficult structure theorem in Theorem 1.5.

² Unfortunately, our reduction to this key proposition is somewhat expensive with regards to the quantitative bounds, and is responsible for the double exponential lower bound required on n .

2 The Key Examples

The aim of this section is to describe the various key examples of sets with few ordinary lines or many 3-rich lines. In particular we describe the sets X_{2m} in (1.1) and the Sylvester examples appearing in the various cases of the main structure theorem, Theorem 1.5. We will also mention some important “near-counterexamples” which do not actually exist as finite counterexamples to the structure theorem, but nevertheless are close enough to genuine counterexamples that some attention must be given in the analysis to explicitly exclude variants of these examples from the list of possible configurations. All of these examples are connected to the group (or pseudo-group) structure on a cubic curve, or equivalently to Chasles’s version of the Cayley–Bacharach theorem as described in Proposition 4.1 below. The main variation in the examples comes from the nature of the cubic curve being considered, which may or may not be irreducible and/or nonsingular.

Böröczky and near-Böröczky examples We begin with the sets X_{2m} from (1.1), together with some slight perturbations of these sets described by Böröczky (as cited in Ref. [10]). These sets, it turns out, provide the examples of non-collinear sets of n points with the fewest number of ordinary lines, at least for n large.

Proposition 2.1 (Böröczky examples) *Let $m \geq 3$ be an integer. Then we have the following.*

- (i) *The set X_{2m} contains $2m$ points and spans precisely m ordinary lines.*
- (ii) *The set X_{4m} together with the origin $[0, 0, 1]$ contains $4m + 1$ points and spans precisely $3m$ ordinary lines.*
- (iii) *The set X_{4m} minus the point $[0, 1, 0]$ on the line at infinity contains $4m - 1$ points and spans precisely $3m - 3$ ordinary lines.*
- (iv) *The set X_{4m+2} minus any of the $2m + 1$ points on the line at infinity contains $4m + 1$ points and spans $3m$ ordinary lines.*

Thus, if we define a function $f : \mathbb{N} \rightarrow \mathbb{N}$ by setting $f(2m) := m$, $f(4m + 1) := 3m$ and $f(4m - 1) := 3m - 3$, then there is an example of a set of n points in \mathbb{RP}^2 , not all on a line, spanning $f(n)$ ordinary lines.

Proof This is a rather straightforward check, especially once one has drawn suitable pictures. Whilst the unit circle together with the line at infinity form a pleasant context for calculational work, drawing configurations involving the line at infinity is problematic. In the four diagrams below, Figs. 4, 5, 6 and 7, we have applied a projective transformation to aid visualisation. First of all we applied a rotation about the origin through $\pi/12$, and followed this by the projective map $[x, y, z] \mapsto [-y, x, 2z + x]$. The unit circle is then sent to the ellipse whose equation in the affine plane is $4x^2 + 3(y + \frac{1}{3})^2 = \frac{4}{3}$, while the line at infinity is sent to the horizontal line $y = 1$. The origin is mapped to itself, and the point $[0, 1, 0]$ at infinity now has coordinates $(-\cot(\pi/12), 1) \approx (-3.73, 1)$. In the pictures, ordinary lines are red and lines with three or more points of P are dotted green.

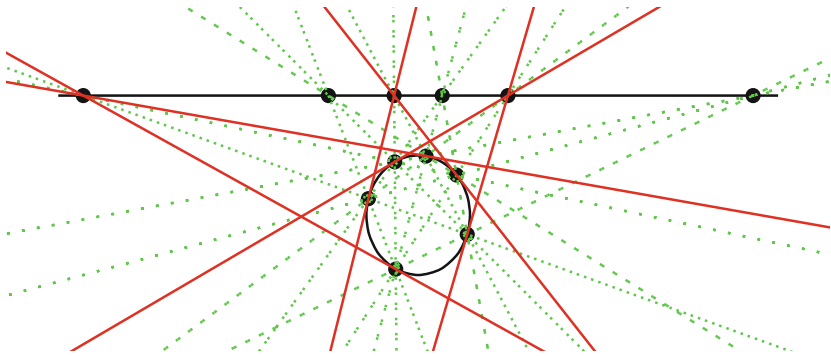


Fig. 4 The Böröczky example X_{12} , a set with $n = 12$ points and 6 ordinary lines. The ordinary lines (in red) are just the tangent lines to the 6th roots of unity on the unit circle (Color figure online)

It is helpful to note that the line joining

$$\left[\cos \frac{2\pi j}{m}, \sin \frac{2\pi j}{m}, 1 \right]$$

and

$$\left[\cos \frac{2\pi j'}{m}, \sin \frac{2\pi j'}{m}, 1 \right]$$

passes through the point

$$\left[-\sin \frac{\pi(j + j')}{m}, \cos \frac{\pi(j + j')}{m}, 0 \right]$$

on the line at infinity (cf. the proof of Proposition 7.3).

For case (i), the ordinary lines are the m tangent lines to the m th roots of unity. The case $m = 6$ is depicted in Fig. 4.

In case (ii), the ordinary lines are the $2m$ tangent lines to the $2m$ th roots of unity together with the m lines joining the origin $[0, 0, 1]$ to $[-\sin \frac{\pi j}{2m}, \cos \frac{\pi j}{2m}, 0]$, j even. The case $m = 6$ is depicted in Fig. 5.

In case (iii), the ordinary lines are the $2m$ tangent lines to the $2m$ th root of unity *except* for those at the points $[\pm 1, 0, 1]$, whose corresponding point at infinity has now been removed. However we do have $m - 1$ new ordinary lines, the vertical lines joining $[\cos \frac{\pi j}{m}, \sin \frac{\pi j}{m}, 1]$ and $[\cos \frac{\pi(2m-j)}{m}, \sin \frac{\pi(2j-m)}{m}, 1]$ for $j = 1, \dots, m - 1$. The case $m = 6$ is illustrated in Fig. 6.

Finally, in case (iv) the ordinary lines are the $2m + 1$ tangent lines to the $(2m + 1)$ th roots of unity *except* for one whose corresponding point at infinity has been removed, together with m new ordinary lines joining pairs of roots of unity. This is illustrated in Fig. 7 in the case $m = 2$. □

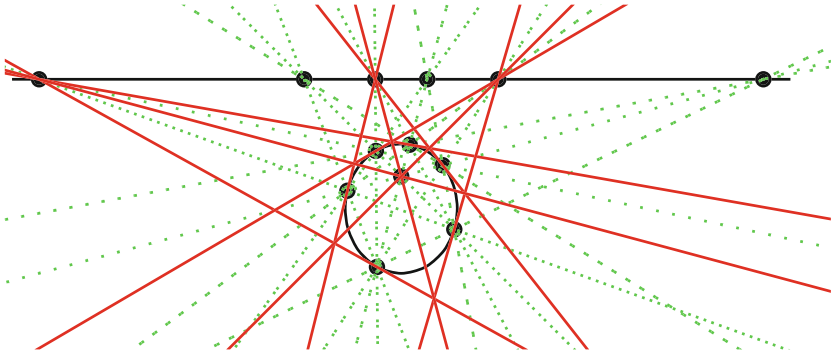


Fig. 5 The Böröczky example X_{12} together with the origin $[0, 0, 1]$, a set with $n = 13$ points and 9 ordinary lines. The ordinary lines (in red) are just the tangent lines to the 6th roots of unity on the unit circle, plus 3 extra lines through the origin and 3 of the points on the line at infinity (Color figure online)

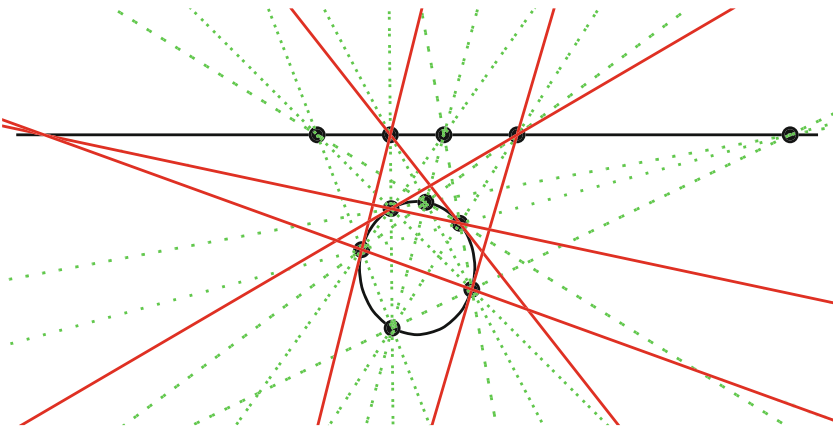


Fig. 6 The Böröczky example X_{12} minus the point at infinity $[0, 1, 0]$, a set with $n = 11$ points and 6 ordinary lines. The ordinary lines (in red) are the 4 tangent lines to the 6th roots of unity on the unit circle not through the point at infinity, plus 2 extra lines passing through the point at infinity (Color figure online)

We remark that Proposition 2.1 illustrates a basic fact, namely that if one adds or removes K points to an n -point configuration, then the number of ordinary lines (or 3-rich lines) is modified by at most $O(Kn + K^2)$; this can be seen by first considering the $K = 1$ case and then iterating. This stability with respect to addition or deletion of a few points is reflected in the conclusions of the various structural theorems in this paper.

We may now state our more precise version of the Dirac–Motzkin conjecture for large n .

Theorem 2.2 (Sharp threshold for Dirac–Motzkin) *Let the function $f : \mathbb{N} \rightarrow \mathbb{N}$ be defined by setting $f(2m) := m$, $f(4m + 1) := 3m$ and $f(4m - 1) := 3m - 3$. There is an n_0 such that the following is true. If $n \geq n_0$ and if P is a set of n points in \mathbb{RP}^2 , not all on a line, then P spans at least $f(n)$ ordinary lines. Furthermore if equality*

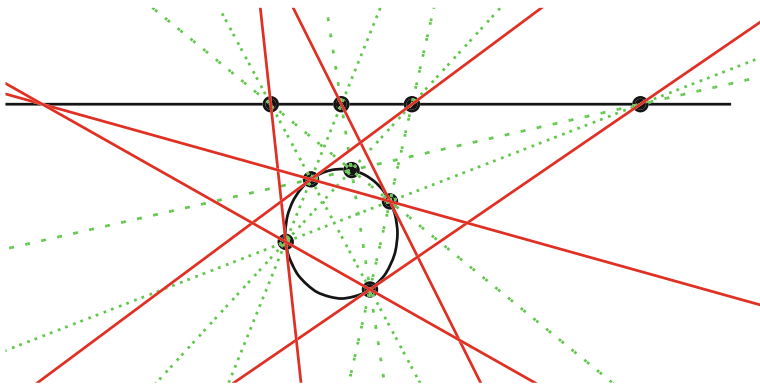


Fig. 7 The Böröczky example X_{10} minus the point at infinity $[0, 1, 0]$, a set with $n = 9$ points and 6 ordinary lines. The ordinary lines (in red) are the 4 tangent lines to the 6th roots of unity on the unit circle not through the point at infinity, plus 2 extra lines passing through the point at infinity (Color figure online)

occurs then, up to a projective transformation, P is one of the Böröczky examples described in Proposition 2.1 above.

Remark Note in particular that there is an essentially unique extremal example unless $n \equiv 1 \pmod{4}$, in which case there are two, namely examples (ii) and (iv) above. Note that, all of the examples in (iv) are equivalent up to rotation.

Let us record, in addition to the Böröczky examples mentioned in Proposition 2.1, the following near-extremal example.

Proposition 2.3 (Near-Böröczky example) *The set X_{4m} minus the point $[-\sin \frac{\pi}{2m}, \cos \frac{\pi}{2m}, 0]$ on the line at infinity contains $4m - 1$ points and spans $3m$ ordinary lines.*

Proof This is illustrated in Fig. 8 in the case $m = 3$. The ordinary lines are the $2m$ tangent lines to the $2m$ th roots of unity as well as m lines joining $[\cos \frac{\pi j}{m}, \sin \frac{\pi j}{m}, 1]$ and $[\cos \frac{\pi j'}{m}, \sin \frac{\pi j'}{m}, 1]$ with $j + j' \equiv 1 \pmod{2m}$.

We may now state a still more precise result, which asserts that all configurations not equivalent to one of the above examples must necessarily have a significantly larger number of ordinary lines than $f(n)$, when n is large. In fact there must be at least $n - O(1)$ ordinary lines in such cases.

Theorem 2.4 (Strong Dirac–Motzkin conjecture) *There is an absolute constant C such that the following is true. If P is a set of n points in \mathbb{RP}^2 , not all on a line, spanning no more than $n - C$ ordinary lines then P is equivalent under a projective transformation to one of the Böröczky examples or to a near-Böröczky example.*

The threshold $n - C$ is sharp except for the constant C . Indeed we will shortly see that finite subgroups of elliptic curves of cardinality n give examples of sets with $n - O(1)$ lines. This gives infinitely many new examples of sets with few ordinary lines which are inequivalent under projective transformation due to the projective invariance of the discriminant of an elliptic curve.

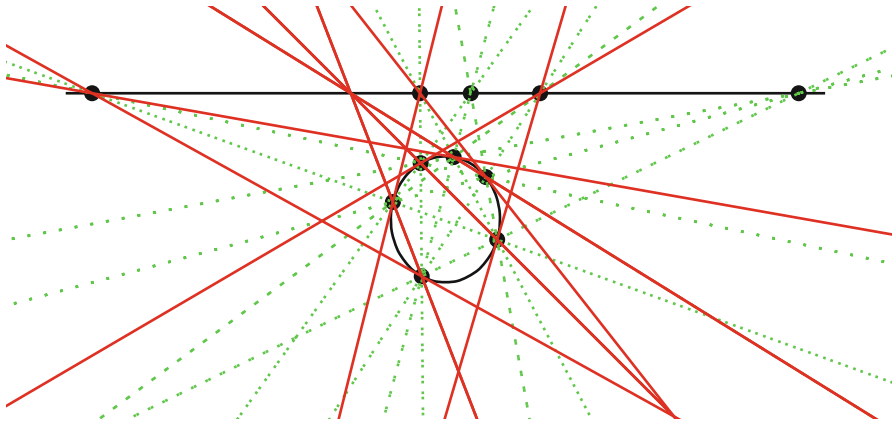


Fig. 8 The near-Böröczky example with $n = 11$ points and 9 ordinary lines. The ordinary lines (in red) are the 6 tangent lines to the 6th roots of unity on the unit circle plus 3 lines passing through the removed point $[-\sin \frac{\pi}{6}, \cos \frac{\pi}{6}, 0]$ (Color figure online)

Sylvester’s cubic curve examples We turn now to Sylvester’s examples of point sets coming from cubic curves, as further discussed by Burr et al. [5]. While these do not provide the best examples of sets with few ordinary lines, it appears that consideration of them is essential in order to solve the Dirac–Motzkin problem. Of course, they also feature in the statement of our main structural result, Theorem 1.5, and are optimal for the orchard problem (see Sect. 9). Finally, they provide essentially different examples of sets with $n + O(1)$ ordinary lines to any of those considered so far.

For a leisurely discussion of all the projective algebraic geometry required in this paper, including an extensive discussion of cubic curves, we recommend the book [3].

Let γ be any irreducible cubic curve. It is known (see [3, Chap. 12]) that γ has a point of inflection, that is to say a point where the tangent meets γ to order 3. By moving this to the point $[0, 1, 0]$ at infinity, we may bring γ into the form $y^2 = f(x)$ in affine coordinates, where $f(x)$ is a cubic polynomial. If γ is smooth then it is called an *elliptic curve*. An elliptic curve may have one or two components; these two cases are illustrated in Fig. 10. If γ has a singular point then it may be transformed into one of the following three (affine) forms:

- (nodal case) $y^2 = x^2(x + 1)$;
- (cuspidal case) $y^2 = x^3$;
- (acnodal case) $y^2 = x^2(x - 1)$.

See [3, Theorem 8.3] for details. These three singular cases are illustrated in Figure 9.

We remark that the classification of cubic curves over \mathbb{R} has a long and honourable history dating back to Isaac Newton.

The group law Suppose that γ is an irreducible cubic curve, and write γ^* for the set of nonsingular points of γ . If γ is smooth then of course $\gamma = \gamma^*$, and in this case γ is an elliptic curve. We may define an abelian group structure on γ^* by taking the identity O to be a point of inflection on γ^* and, roughly speaking, $P \oplus Q \oplus R = O$ if and only if P, Q, R are collinear. The “roughly speaking” refers to the fact that we must take appropriate account of multiplicity, thus $P \oplus P \oplus Q = O$ if the tangent

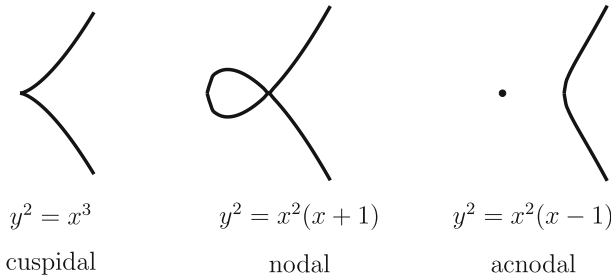


Fig. 9 The three different types of singular cubic curve

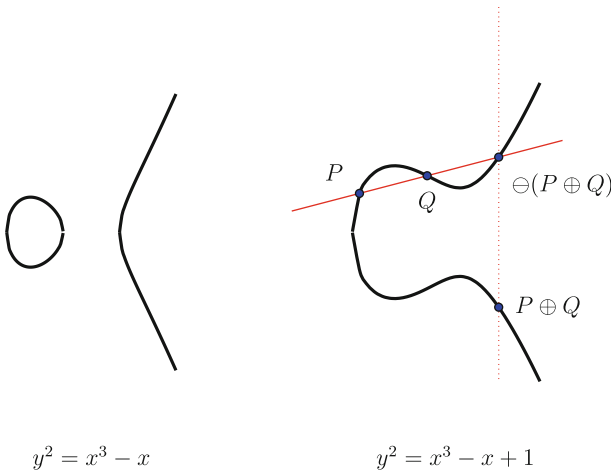


Fig. 10 Two elliptic curves, illustrating the group law and showing the two possibilities for the group structure

to γ at P also passes through Q . The inverse $\ominus P$ of P is defined using the fact that $\ominus P, O$ and P are collinear. See [3, Chap. 9] for more details, including a proof that this does indeed give γ^* the structure of an abelian group.

We have the following theorem regarding the nature of γ^* as a group.

Theorem 2.5 *Let γ be an irreducible cubic curve, and let γ^* be the set of its nonsingular points. Then we have the following possibilities for γ^* , considered as a group:*

- (elliptic curve case) \mathbb{R}/\mathbb{Z} or $\mathbb{R}/\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, depending on whether γ has 1 or 2 connected components;
- (nodal case) $\mathbb{R} \times \mathbb{Z}/2\mathbb{Z}$;
- (cuspidal case) \mathbb{R} ;
- (acnodal case) \mathbb{R}/\mathbb{Z} .

Once again, details may be found in [3]. Thinking about the curves topologically, the theorem is reasonably evident. In the three singular cases isomorphisms $\phi : G \rightarrow \gamma^*$ can be given quite explicitly, as detailed in the following list.

- In the nodal case $y^2 = x^2(x + 1)$, the map $\phi : \mathbb{R} \times \mathbb{Z}/2\mathbb{Z} \rightarrow \gamma^*$ defined by $\phi(x, \varepsilon) = (t^2 - 1, t(t^2 - 1))$, where $t = \coth x$ if $\varepsilon = 0$ and $t = \tanh x$ if $\varepsilon = 1$ provides an isomorphism;
- In the cuspidal case $y^2 = x^3$, the map $\phi : \mathbb{R} \rightarrow \gamma^*$ defined by $\phi(x) = (\frac{1}{x^3}, \frac{1}{x^2})$ provides an isomorphism;
- In the acnodal case $y^2 = x^2(x - 1)$, the map $\phi : \mathbb{R}/\mathbb{Z} \rightarrow \gamma^*$ defined by $\phi(x) = (t^2 + 1, t(t^2 + 1))$, where $t = \cot(\pi x)$, provides an isomorphism.

We leave the reader to provide the details. In the nodal case (for example) we recommend first proving that $(t^2 - 1, t(t^2 - 1), (u^2 - 1, u(u^2 - 1))$ and $(v^2 - 1, v(v^2 - 1))$ are collinear if and only if $-v = (1 + tu)/(t + u)$.

The following maps in the other direction, described in Silverman [30, III. 7] (the acnodal case is described in [30, Exercise 3.15]) are perhaps even tidier. Here $\infty = [0, 1, 0]$.

- In the nodal case the map $(x, y) \mapsto (y - x)/(y + x)$ and $\infty \mapsto 1$ gives an isomorphism from γ^* to $\mathbb{R}^* \cong \mathbb{R} \times \mathbb{Z}/2\mathbb{Z}$;
- In the cuspidal case the map $(x, y) \mapsto x/y$ and $\infty \mapsto 1$ gives an isomorphism from γ^* to \mathbb{R} ;
- In the acnodal case the map $(x, y) \mapsto -(x + iy)^2/x^3$ and $\infty \mapsto 1$ gives an isomorphism from γ^* to the unit circle S^1 in the complex plane.

Sylvester’s examples By a *Sylvester example* E_n we mean a set of n points P in the plane which corresponds to a subgroup of order n of an irreducible cubic curve γ . If $n > 2$ the existence of such an example requires γ to be either an elliptic curve or an acnodal³ cubic curve, by the classification of the group structure of γ described in Theorem 2.5. A Sylvester example coming from an elliptic curve is depicted in Fig. 11.

As it turns out, Sylvester examples have somewhat more ordinary lines than the Böröczky examples, namely $n + O(1)$ instead of $n/2 + O(1)$ or $3n/4 + O(1)$, and are thus not extremisers for the Dirac–Motzkin conjecture. However, due to the more evenly distributed nature of the Sylvester examples, they have significantly more 3-rich lines. Indeed, the following is essentially established in Ref. [5].

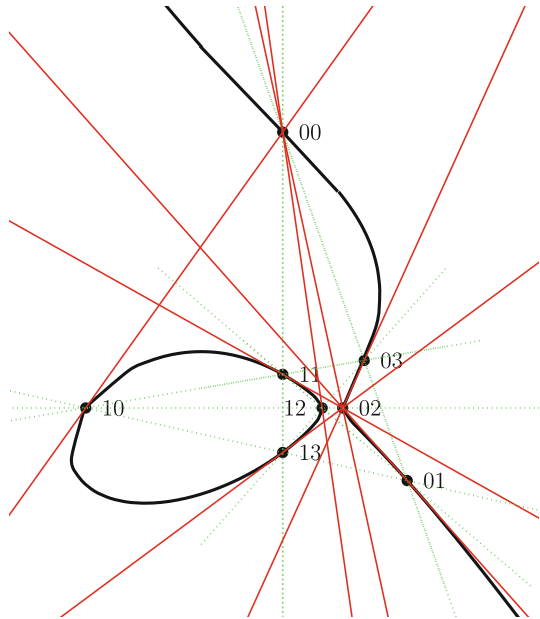
Proposition 2.6 *Let $n \geq 3$, and let E_n be a subgroup of order n in γ^* , the group of nonsingular points of an irreducible cubic curve γ (which must be an elliptic curve or an acnodal cubic). Then E_n spans $n - 1 - 2 \cdot \mathbf{1}_{3|n}$ ordinary lines and $\lfloor \frac{n(n-3)}{6} \rfloor + 1$ 3-rich lines, where $\mathbf{1}_{3|n}$ is equal to 1 when 3 divides n and zero otherwise. Furthermore, if $x \in E$ is such that $x \notin E_n$ and $x \oplus x \oplus x \in E_n$ then $E_n \oplus x$ has $n - 1$ ordinary lines and $\lfloor \frac{n(n-3)}{6} \rfloor$ 3-rich lines.*

Proof Let N_2 be the number of ordinary lines, and N_3 be the number of 3-rich lines. From Bézout’s theorem no line can meet E_n in more than three points, and so by double counting we have the identity

$$N_2 + \binom{3}{2}N_3 = \binom{n}{2}.$$

³ We do not know whether Sylvester himself was interested in the acnodal case. We thank Frank de Zeeuw for correcting an oversight in the first version of this paper by drawing it to our attention.

Fig. 11 A Sylvester example with $n = 8$, the subgroup being isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$. The labels reflect the group structure, thus 03 corresponds to the element $(0, 3) \in \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$. This comes from an elliptic curve with equation $y^2 = x^3 - \frac{1}{36}x^2 - \frac{5}{36}x + \frac{25}{1296} = 0$ to which we have applied the projective transformation $[x, y, z] \mapsto [x, y, x + y + z]$, so that the point at infinity maps to the point $(0, 1)$ in the affine plane (which is then an inflection point for the curve). There are 7 ordinary lines, marked in red, and also 7 3-rich lines, marked in dotted green (Color figure online)



A brief computation (splitting into three cases depending on the residue of n modulo 3) then shows that $N_3 = \lfloor \frac{n(n-3)}{6} \rfloor + 1$ if and only if $N_2 = n - 1 - 2 \cdot \mathbf{1}_{3|n}$. But from the group law the number of ordinary lines is precisely equal to the number of elements $a \in E_n$ such that $-2a$ is distinct from a , or in other words the number n of elements in E_n minus the number of third roots in E_n . But γ^* is isomorphic as a group to either \mathbb{R}/\mathbb{Z} or $(\mathbb{R}/\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$, and so E_n is isomorphic to either $\mathbb{Z}/n\mathbb{Z}$ or to $(\mathbb{Z}/(n/2)\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. It has $1 + 2 \cdot \mathbf{1}_{3|n}$ third roots in either case, and the claim follows.

The analysis in the shifted case $E_n \oplus x$ is analogous, the only difference being that $E_n \oplus x$ does not contain any third roots of unity. □

Remarks For the sake of comparison, the n -point examples in Proposition 2.1 can all be computed to have $n^2/8 + O(n)$ 3-rich lines instead of $n^2/6 + O(n)$ for the Sylvester examples. This discrepancy can be explained by the existence of a high-multiplicity line with $n/2 + O(1)$ points in those examples. This absorbs many of the pairs of points that could otherwise be generating 3-rich lines.

We note also that the acnodal case allows for a quite explicit construction of a set of n points defining $\sim n^2/6$ 3-rich lines, without the use of the Weierstrass \wp -function which would be necessary in the elliptic curve case. We leave the reader to supply the details, using the parametrisation detailed after the statement of Theorem 2.5. We are not sure whether this point has been raised in the literature before.

Near-counterexamples In addition to the actual examples coming from Böröczky’s constructions and from elliptic curve subgroups, there are also some important “near-counterexamples” which do not *directly* enter into the analysis (because they involve an infinite number of points, rather than a finite number), but which nevertheless appear

to indirectly complicate the analysis by potentially generating spurious counterexamples to the structural theory of points with few ordinary lines. These then need to be eliminated by additional arguments.

As with the previously discussed examples, the near-counterexamples discussed here will lie on cubic curves. But whilst the actual examples were on an elliptic curve, an acnodal singular cubic curve, or on the union of a conic and a line, the near-counterexamples will lie on three lines (which may or may not be concurrent), or on a non-acnodal singular cubic curve.

We first consider a near-counterexample on three concurrent lines. Up to projective transformation, one can take the lines to be the parallel lines

$$\begin{aligned} \ell_1 &:= \{[x_1, 0, 1] : x_1 \in \mathbb{R}\} \cup \{[1, 0, 0]\}, \\ \ell_2 &:= \{[x_2, 1, 1] : x_2 \in \mathbb{R}\} \cup \{[1, 0, 0]\}, \\ \ell_3 &:= \{[x_3, 2, 1] : x_3 \in \mathbb{R}\} \cup \{[1, 0, 0]\}. \end{aligned}$$

Observe that $[x_1, 0, 1]$, $[x_2, 1, 1]$ and $[x_3, 2, 1]$ are colinear if and only if $x_1 + x_3 = 2x_2$. Thus, if we consider the infinite point set

$$\begin{aligned} P &:= \{[n_1, 0, 1] : n_1 \in \mathbb{Z}\} \cup \{[n_2, 1, 1] : n_2 \in \frac{1}{2}\mathbb{Z}\} \\ &\cup \{[n_3, 2, 1] : n_3 \in \mathbb{Z}\} \end{aligned} \tag{2.1}$$

then there are no ordinary lines whatsoever; every line joining a point in $P \cap \ell_1$ with a point in $P \cap \ell_2$ meets a point in $P \cap \ell_3$, and similarly for permutations. If \mathbb{Z} could somehow have a non-trivial finite subgroup, then one could truncate this example into a counterexample to the Sylvester–Gallai theorem, i.e. a finite set with no ordinary lines. Of course, this cannot actually happen, but this example strongly suggests that one needs to somehow use the torsion-free nature of the additive group \mathbb{R} at some point in the arguments, for instance by exploiting arguments based on convexity, or by using additive combinatorics results exploiting the ordered nature of \mathbb{R} . One such example, a variant of which we prove in Lemma A.3, is the trivial inequality $|A + B| \geq |A| + |B| - 1$ for finite subsets A, B of \mathbb{R} . This can be viewed as a quantitative version of the assertion that \mathbb{R} has no non-trivial finite subgroups.

There is a similar near-counterexample involving three non-concurrent lines. Again, after applying a projective transformation, we may work with the lines

$$\begin{aligned} \ell_1 &:= \{[x, 0, 1] : x \in \mathbb{R}\} \cup \{[1, 0, 0]\}, \\ \ell_2 &:= \{[0, y, 1] : y \in \mathbb{R}\} \cup \{[0, 1, 0]\}, \\ \ell_3 &:= \{[-z, 1, 0] : z \in \mathbb{R}\} \cup \{[1, 0, 0]\}. \end{aligned}$$

Observe that if $x, y, z \in \mathbb{R}^\times := \mathbb{R} \setminus \{0\}$, then $[x, 0, 1]$, $[0, y, 1]$ and $[-z, 1, 0]$ are concurrent precisely when $z = x/y$. Thus, if we consider the infinite point set

$$\begin{aligned} P &:= \{[2^{n_1}, 0, 1] : n_1 \in \mathbb{Z}\} \cup \{[0, 2^{n_2}, 1] : n_2 \in \mathbb{Z}\} \\ &\cup \{[-2^{n_3}, 1, 0] : n_3 \in \mathbb{Z}\}, \end{aligned} \tag{2.2}$$

then again there are no ordinary lines: every line joining a point in $P \cap \ell_1$ with a point in $P \cap \ell_2$ meets a point in $P \cap \ell_3$, and similarly for permutations. As before, this example suggests that the (essentially) torsion-free nature of the multiplicative group \mathbb{R}^\times must somehow come into play at some point in the argument.

Finally, we give an example that lies on a cuspidal singular cubic curve, which after projective transformation can be written as

$$\gamma := \{[x, y, z] : yz^2 = x^3\}.$$

Removing the singular point at $[0, 1, 0]$, we may parameterise the smooth points γ^* of this curve by $\{[t, t^3, 1] : t \in \mathbb{R}\}$. One can compute after a brief determinant computation that three distinct smooth points $[t_1, t_1^3, 1]$, $[t_2, t_2^3, 1]$ and $[t_3, t_3^3, 1]$ on the curve are concurrent precisely when $t_1 + t_2 + t_3 = 0$. Thus, if one sets P to be the infinite set

$$P := \{[n, n^3, 1] : n \in \mathbb{Z}\} \tag{2.3}$$

then there are very few ordinary lines—indeed only those lines that are tangent to γ at one point $[n, n^3, 1]$ and meet γ at a second point $[(-2n), (-2n)^3, 1]$ for some $n \in \mathbb{Z} \setminus \{0\}$ will be ordinary. This example can be viewed as a degenerate limit of the Sylvester examples E_n when the discriminant is sent to zero and n sent to infinity. Again, finitary versions of this example can be ruled out, but only after one exploits the torsion-free nature of the group associated to γ^* , which in this case is isomorphic to \mathbb{R} . Similar remarks also apply to nodal singular cubic curves such as $\{[x, y, z] : y^2z = x^3 + x^2z\}$, the smooth points of which form a group isomorphic to $\mathbb{R} \times (\mathbb{Z}/2\mathbb{Z})$, which is essentially torsion free in the sense that there are no large finite subgroups.

A variant of the example (2.3) lies on the union

$$\{[x, y, z] : yz = x^2\} \cup \{[x, y, z] : z = 0\}$$

of a parabola and the line at infinity. Observe that two points $[t_1, t_1^2, 1]$, $[t_2, t_2^2, 1]$ on the parabola and a point $[0, t_3, 1]$ on the line at infinity with $t_1, t_2, t_3 \in \mathbb{R}$ are concurrent if and only if $t_3 = t_1 + t_2$. Thus, the infinite set

$$P := \{[n, n^2, 1] : n \in \mathbb{Z}\} \cup \{[0, n, 1] : n \in \mathbb{Z}\}, \tag{2.4}$$

which can be viewed as a degenerate limit of a Böröczky example, has very few ordinary lines (namely, the line tangent to the parabola at one point $[n, n^2, 1]$ and also passing through $[0, 2n, 1]$).

The existence of these near-counterexamples forces us to use a somewhat ad hoc case-by-case analysis. Tools such as Chasles’s version of the Cayley–Bacharach theorem, which are valid for all cubic curves, get us only so far. They must be followed up by more specialised arguments exploiting the torsion or lack thereof in the group structure. In this way we can rule out near-counterexamples involving triples of lines, or singular irreducible cubics, until only the Böröczky and Sylvester type of examples remain.

3 Melchior’s Proof of the Sylvester–Gallai Theorem

In this section, we review Melchior’s proof [25] of the Sylvester–Gallai theorem. As mentioned in the Sect. 1, this is the starting point for all of our arguments.

Theorem 1.1 (Sylvester–Gallai, again) *Suppose that P is a finite set of points in the plane, not all on one line. Then P spans at least one ordinary line.*

Proof Let P be a set of n points in \mathbb{RP}^2 . Consider the dual collection $P^* := \{p^* : p \in P\}$ of n lines in \mathbb{RP}^2 . These lines determine a graph⁴ Γ_P in \mathbb{RP}^2 whose vertices are the intersections of pairs of lines p_1^*, p_2^* (or equivalently points ℓ^* , where ℓ is a line joining two or more points of P), and whose edges are (projective) line segments of lines in P^* connecting two vertices of Γ_P with no vertex in the interior. Note that as the points in P were assumed not to lie on one line, every line in P^* must meet at least two vertices of Γ_P ; in particular, the graph Γ_P contains no loops. (It is however possible for a line to meet exactly two vertices in Γ_P , in which case those two vertices are joined by two edges, rather than one.) Also, by construction, each vertex of Γ_P is incident to at least two lines in P^* . As such, the graph Γ_P partitions the projective plane \mathbb{RP}^2 into some number V of vertices, some number E of edges, and some number F of faces, each of which is the projective image of a polygon. In particular, each face has at least three edges, and any edge is incident to two distinct faces.

By Euler’s formula in the projective plane \mathbb{RP}^2 we have⁵

$$V - E + F = 1. \tag{3.1}$$

To proceed further, suppose that for each $k = 2, 3, 4, \dots$ the set P has N_k lines containing precisely k points of P . Then V , which by duality is the number of lines defined by pairs of points in P , satisfies

$$V = \sum_{k=2}^n N_k. \tag{3.2}$$

Furthermore the degree $d(\ell^*)$ of a vertex ℓ^* in our graph is twice the number of lines in P^* passing through ℓ^* , which is $2|P \cap \ell|$. Thus, summing over all lines ℓ ,

$$2E = \sum_{\ell} d(\ell^*) = 2 \sum_{\ell} |P \cap \ell| = \sum_{k=2}^n 2kN_k. \tag{3.3}$$

Finally, for $s = 3, 4, 5, \dots$ write M_s for the number of faces in Γ_P with s edges. Since each edge is incident to exactly two faces, we have

⁴ Strictly speaking, Γ_P determines a *drawing* of a graph in the projective plane, because we are viewing the edges as curves in \mathbb{RP}^2 rather than abstract pairs of vertices, but we shall abuse notation by identifying a graph with its drawing.

⁵ The Euler characteristic $\chi(\mathbb{RP}^2) = 1$ of the projective plane is of course half of the Euler characteristic $\chi(S^2) = 2$ of the sphere, as the latter is a double cover of the former.

$$2E = \sum_{s=3}^n sM_s. \tag{3.4}$$

Combining (3.1), (3.2), (3.3) and (3.4) gives the following expression for N_2 , the number of ordinary lines:

$$N_2 = 3 + \sum_{k=4}^n (k - 3)N_k + \sum_{s=4}^n (s - 3)M_s. \tag{3.5}$$

It follows immediately that $N_2 \geq 3$, which of course implies the Sylvester-Gallai theorem. □

After discarding the non-negative term $\sum_{s=4}^n (s - 3)M_s$, Eq. (3.5) implies *Melchior’s inequality*

$$N_2 \geq 3 + \sum_{k=4}^n (k - 3)N_k.$$

In this paper, however, we will need to save the term $\sum_{s=4}^n (s - 3)M_s$, as it gives crucial control on the geometry of the dual configuration Γ_P , ensuring that this configuration resembles a triangulation when N_2 is small. More precisely, we have the following proposition.

Proposition 3.1 (Few bad edges) *Suppose that P is a set of n points in the projective plane \mathbb{RP}^2 , not all on a line, and suppose that P has at most Kn ordinary lines. Consider the planar graph Γ_P obtained by dualising P as described above. Then Γ_P is an “almost triangulation” in the following sense. Say that an edge of Γ_P is good if both of its vertices have degree 6, and if both faces it adjoins are triangles. Say that it is bad otherwise. Then the number of bad edges in Γ_P is at most $16Kn$.*

Proof From (3.5) we have

$$\sum_{s=4}^n sM_s \leq 4 \sum_{s=4}^n (s - 3)M_s \leq 4N_2 \leq 4Kn. \tag{3.6}$$

Secondly, let us observe that

$$\sum_{\ell:d(\ell^*) \neq 6} d(\ell^*) \leq 12Kn. \tag{3.7}$$

To see this, recall that $d(\ell^*) = 2|P \cap \ell|$. We thus obtain

$$\begin{aligned} \sum_{\ell:d(\ell^*) > 6} d(\ell^*) &= 2 \sum_{\ell:|P \cap \ell| > 3} |P \cap \ell| \\ &= \sum_{k \geq 4} 2kN_k \leq 8 \sum_{k \geq 4} (k - 3)N_k \leq 8Kn. \end{aligned}$$

Noting also that

$$\sum_{\ell:d(\ell^*)=4} d(\ell^*) = 2 \sum_{\ell:|P \cap \ell|=2} |P \cap \ell| = 4N_2 \leq 4Kn,$$

(3.7) follows.

Now we can place an upper bound on the number B of bad edges. Each face with $s > 3$ sides gives s bad edges, and each vertex ℓ^* with degree $d(\ell^*) \neq 6$ gives $d(\ell^*)$ bad edges. As these are the only sources of bad edges, we have

$$B \leq \sum_{s>3} sM_s + \sum_{\ell:d(\ell^*)\neq 6} d(\ell^*) \leq 16Kn,$$

by (3.6) and (3.7). □

4 Triangular Structure in the Dual and Cubic Curves

In general, the number of edges overall in Γ_P is expected to be of the order of n^2 (cf. Beck’s theorem [2]). Thus, when K is small, Proposition 3.1 should be viewed as an assertion that almost all the edges of Γ_P are good. For instance, it shows that any dual line p^* , $p \in P$ should contain at most $O(K)$ bad edges on the average. Intuitively, this suggests that Γ_P is an “almost triangulation” in which most vertices have degree 6 and most faces are triangles. In Sect. 5 we will use this information to put the points of P on a small number of cubic curves, which will be our starting point for more powerful structural theorems on P .

By a *cubic curve* we mean a set of points in \mathbb{RP}^2 of the form

$$\{[X, Y, Z] : a_1X^3 + a_2X^2Y + a_3XY^2 + a_4Y^3 + a_5X^2Z + a_6XYZ + a_7Y^2Z + a_8XZ^2 + a_9YZ^2 + a_{10}Z^3 = 0\}$$

for some coefficients $a_1, \dots, a_{10} \in \mathbb{R}$, not all zero, or in other words the locus of a nontrivial homogeneous polynomial of degree 3. Note that, we do not assume this polynomial to be irreducible. In particular, we consider the union of three lines, as well as the union of a conic and a line, to be examples of cubic curves.

A key observation in our arguments will be the fact that pockets of true triangular structure in the dual Γ_P signify a collection of points of P lying on a single cubic curve. Results of this type may be found in Lemmas 4.4 and 4.5 below. A key ingredient will be the following incredibly classical fact from projective geometry, usually known as the *Cayley–Bacharach theorem* (although the case we require was proven by Chasles [9], prior to the more general results of Cayley [8] and Bacharach [1]).

Proposition 4.1 (Chasles) *Suppose that two sets of three lines define nine distinct points of intersection in \mathbb{RP}^2 . Then any cubic curve passing through eight of these points also passes through the ninth.*

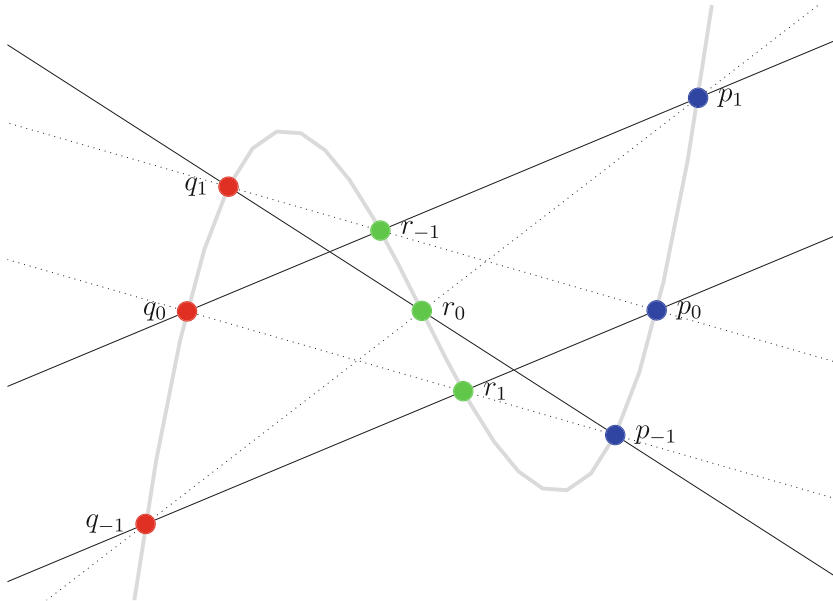


Fig. 12 Chasles’s theorem. There are two sets of three lines: the *solid lines* $\{p_0, q_{-1}, r_1\}, \{p_1, q_0, r_{-1}\}, \{p_{-1}, q_1, r_0\}$ and the *dotted lines* $\{p_0, q_1, r_{-1}\}, \{p_1, q_{-1}, r_0\}$ and $\{p_{-1}, q_0, r_1\}$. The nine points of intersection $p_{-1}, p_0, p_1, q_{-1}, q_0, q_1, r_{-1}, r_0$ and r_1 are all distinct. Any cubic curve passing through 8 of these points also passes through the 9th; one such curve is shown

This situation is shown in Fig. 12 below. See [13] or the blog post [34] for a discussion of this result, including its link to Pappus’s theorem, Pascal’s theorem, and the associativity of the group law on an elliptic curve.

Proposition 4.1 allows one to establish some duality relationships⁶ between triangular grids and cubic curves. We first define what we mean by a triangular grid:

Definition 4.2 (Triangular grid) Let I, J, K be three discrete intervals in \mathbb{Z} (thus I takes the form $\{i \in \mathbb{Z} : i_- \leq i \leq i_+\}$ for some integers i_-, i_+ , and similarly for J and K). A *triangular grid* with dimensions I, J, K is a collection of lines $(p_i^*)_{i \in I}, (q_j^*)_{j \in J}, (r_k^*)_{k \in K}$ in \mathbb{RP}^2 , which we will view as duals of not necessarily distinct points p_i, q_j, r_k in \mathbb{RP}^2 , obeying the following axioms:

- (i) If $i \in I, j \in J, k \in K$ are integers with $i + j + k = 0$, then the lines p_i^*, q_j^*, r_k^* are distinct and meet at a point P_{ijk} . Furthermore, this point P_{ijk} is not incident to any line in the grid which is not already identical to one of the lines p_i^*, q_j^*, r_k^* . Thus, for instance, if $i' \in I$ is such that $p_{i'}^* \neq p_i^*, q_j^*, r_k^*$, then $p_{i'}^*$ cannot contain P_{ijk} .

⁶ Let us note once again that rather similar results were obtained earlier, in a completely different context, by Carnicer and Godés [7].

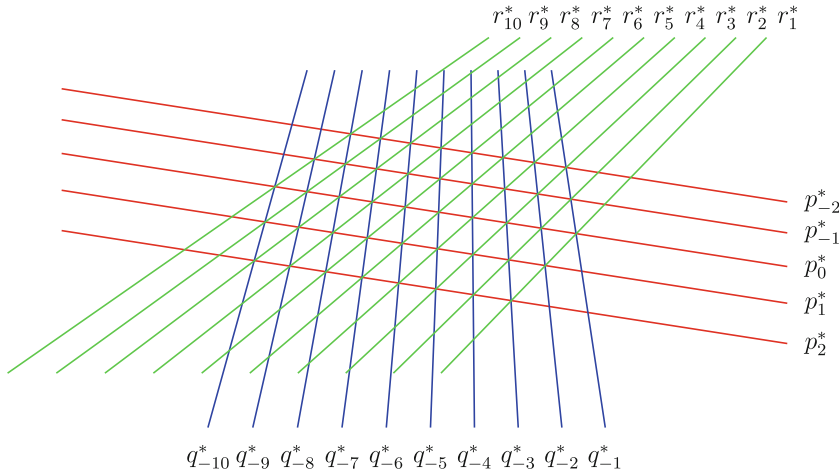


Fig. 13 A triangular grid with dimensions $\{-2, \dots, 2\}$, $\{-10, \dots, -1\}$ and $\{1, \dots, 10\}$

- (ii) If $i \in I, j, j' \in J, k, k' \in K$ are such that $i + j + k = i + j' + k' = 0$ and $0 < |j - j'| \leq 2$ (or equivalently $0 < |k - k'| \leq 2$), then the intersection points P_{ijk} and $P_{ij'k'}$ are distinct. In particular, this forces $q_j^* \neq q_{j'}^*$ and $r_k^* \neq r_{k'}^*$. Similarly for cyclic permutations of i, j, k and of j', k' .

An example of a triangular grid is depicted in Fig. 13.

The following basic consequence of Proposition 4.1 drives our whole argument.

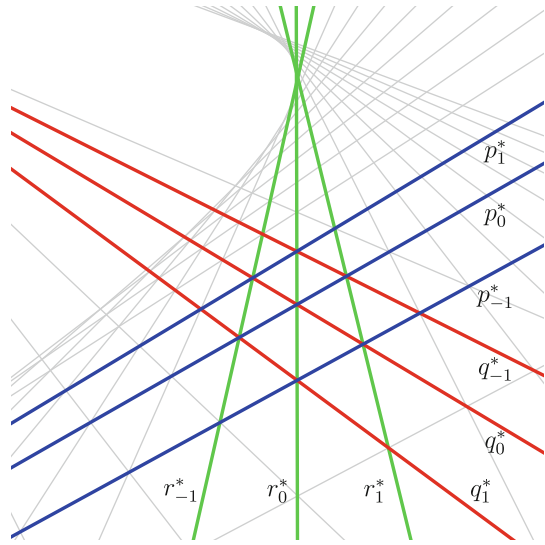
Lemma 4.3 (Completing a hexagon) *Let i_0, j_0, k_0 be integers with $i_0 + j_0 + k_0 = 0$, let $I := \{i_0 - 1, i_0, i_0 + 1\}$, $J := \{j_0 - 1, j_0, j_0 + 1\}$, $K := \{k_0 - 1, k_0, k_0 + 1\}$, and let $(p_i)_{i \in I}, (q_j)_{j \in J}, (r_k)_{k \in K}$ be triples of points whose duals form a triangular grid with dimensions I, J, K . Then the nine points $(p_i)_{i \in I}, (q_j)_{j \in J}, (r_k)_{k \in K}$ are distinct, and any cubic curve which passes through eight of them passes through the ninth.*

Proof By relabeling, we may assume that $i_0 = j_0 = k_0 = 0$, thus the nine points are $p_{-1}, p_0, p_1, q_{-1}, q_0, q_1, r_{-1}, r_0, r_1$. Once it is shown that these nine points are distinct, their duals form a “hexagon” as depicted in Fig. 14 below as part of a larger triangular grid. The configuration in Fig. 14, however, is precisely the dual of the configuration of 9 points appearing in Chasles’s theorem (see Fig. 12), and the claim then follows.

It remains to establish the distinctness of the nine points. By applying Definition 4.2(ii) to the intersections of p_i^*, q_0^*, r_{-i}^* for $i = -1, 0, 1$ we see that the p_{-1}, p_0, p_1 are distinct; similarly for q_{-1}, q_0, q_1 and r_{-1}, r_0, r_1 . Next, from Definition 4.2(i) we see that p_i and q_j are distinct as long as $-1 \leq i + j \leq 1$, and similarly for cyclic permutations. The only remaining claim left to check, up to permutations and reflections, is that p_1 and q_1 are distinct. But if these two points coincided, then the intersections of p_1^*, q_{-1}^*, r_0^* and p_{-1}^*, q_1^*, r_0^* would then also coincide, contradicting Definition 4.2(ii). □

We now iterate the above proposition.

Fig. 14 The dual of the configuration in Chasles’s theorem, showing a “hexagon” formed by the duals of two sets of three lines. The light grey lines are duals of other points on the cubic curve shown in Fig. 12, specifically those points in longer arithmetic progressions (in the group law on γ) containing the p_i, q_j, r_k . We have included them mainly for aesthetic interest, but also as a more complicated example of a triangular grid



Lemma 4.4 *Suppose that $m \geq 4$ is an integer and that i_-, i_+ are integers with $2 \leq i_+ \leq m - 2$ and $2 - m \leq i_- \leq -1$. Suppose that we have a collection of points $(p_i)_{i_-\leq i \leq i_+}, (q_j)_{-m \leq j \leq -1}$ and $(r_k)_{1 \leq k \leq m}$ in $\mathbb{R}P^2$ whose duals form a triangular grid with the indicated dimensions (The case $i_- = -2, i_+ = 2$ and $m = 10$ is illustrated in Fig. 13). Then all of the points p_i, q_j, r_k lie on a single cubic curve γ .*

Proof Consider the nine points $p_{-1}, p_0, p_1, p_2, q_{-3}, q_{-2}, q_{-1}, r_1, r_2$. The space of cubic homogeneous polynomials is a vector space of dimension 10, and so by straightforward linear algebra there is a cubic curve γ containing these nine points $p_{-1}, p_0, p_1, p_2, q_{-3}, q_{-2}, q_{-1}, r_1$ and r_2 . (Note that it is not necessary for the nine points to be distinct in order to obtain this claim.) We will now claim that all the remaining points p_i, q_j, r_k in the configuration also lie on γ .

Firstly, by applying Lemma 4.3 to the set

$$p_{-1}, p_0, p_1, q_{-3}, q_{-2}, q_{-1}, r_1, r_2, r_3$$

we see that as eight of the points already lie in γ , the ninth point r_3 must also. We now know that the 10 points

$$p_{-1}, p_0, p_1, p_2, q_{-3}, q_{-2}, q_{-1}, r_1, r_2, r_3$$

all lie on γ . Now apply Lemma 4.3 to the set

$$p_0, p_1, p_2, q_{-4}, q_{-3}, q_{-2}, r_1, r_2, r_3.$$

We conclude that q_{-4} also lies on γ , so now the 11 points

$$p_{-1}, p_0, p_1, p_2, q_{-4}, q_{-3}, q_{-2}, q_{-1}, r_1, r_2, r_3$$

all lie on γ .

Next apply Lemma 4.3 to the set

$$p_{-1}, p_0, p_1, q_{-4}, q_{-3}, q_{-2}, r_2, r_3, r_4$$

to conclude that r_4 lies on γ . We now know that the 12 points

$$p_{-1}, p_0, p_1, p_2, q_{-4}, q_{-3}, q_{-2}, q_{-1}, r_1, r_2, r_3, r_4$$

all lie on γ .

By shifting the q indices down by one and r indices up by one repeatedly, we may then inductively place q_{-k} and r_k in γ for all $4 \leq k \leq m$. Finally, by applying Lemma 4.3 inductively to the sets

$$p_{i-1}, p_i, p_{i+1}, q_{-i-3}, q_{-i-2}, q_{-i-1}, r_1, r_2, r_3$$

for $i = 2, \dots, i_+ - 1$ (noting that $-i - 3 \geq -i_+ - 2 \geq -m$) we may place p_i in γ for all $2 < i \leq i_+$, and similarly by applying Lemma 4.3 inductively to

$$p_{-i-1}, p_{-i}, p_{-i+1}, q_{-3}, q_{-2}, q_{-1}, r_{i+1}, r_{i+2}, r_{i+3}$$

for $i = 1, \dots, -i_- - 1$ we can also place p_{-i} in γ for all $1 < i \leq -i_-$. This concludes the proof of the claim. □

This lemma is already enough to imply our most basic structural result for sets with few ordinary lines, Proposition 5.1. To get stronger results, such as Proposition 5.3, we need to perform a deeper analysis. The new feature in the following lemma is the last statement.

Lemma 4.5 *Suppose that $L \geq 10$ and that $m \geq 10L$. Suppose that we have a collection of $4L + 1 + 2m$ points $(p_i)_{-2L \leq i \leq 2L}$, $(q_j)_{-m \leq j \leq -1}$ and $(r_k)_{1 \leq k \leq m}$ in \mathbb{RP}^2 whose duals form a $(4L + 1) \times m \times m$ triangular grid with the indicated dimensions. Assume furthermore that the points p_i, q_j, r_k are all distinct. Then all of the points p_i, q_j, r_k lie on a single cubic curve γ , each irreducible component of which contains at least L of the points p_i, q_j, r_k .*

Proof Note from Definition 4.2 and the distinctness of the p_i, q_j, r_k that the intersection points $P_{ijk} = p_i^* \cap q_j^* \cap r_k^*$ in the grid are all distinct.

That all the p_i, q_j, r_k lie on a single cubic curve γ follows from Lemma 4.4.

If γ is already an irreducible cubic then we are done. By enlarging γ via the addition of extra lines if necessary, we may otherwise suppose that we are in one of the following two cases:

- Case 1: γ is the union of three distinct lines ℓ, ℓ', ℓ'' ;
- Case 2: γ is the union of an irreducible conic σ and a line ℓ .

In each case, we are to show that all irreducible components contain at least L points p_i, q_j, r_k .

In Case 1, consider a triple of points p_i, q_j, r_k with $i + j + k = 0$. Since p_i, q_j, r_k are collinear and lie on $\ell \cup \ell' \cup \ell''$, one of the following two possibilities holds:

- (i) One of the three lines ℓ, ℓ', ℓ'' is incident to all three of p_i, q_j, r_k (i.e. the line $\overline{\{p_i, q_j, r_k\}}$ is one of $\ell, \ell',$ or ℓ'');
- (ii) p_i, q_j, r_k lie on one of each of the lines ℓ, ℓ', ℓ'' (for instance, one could have $p_i \in \ell', q_j \in \ell'', r_k \in \ell$, or any of the other five possible permutations). Note that we allow a point to lie on more than one of the lines ℓ, ℓ', ℓ'' .

First of all note that (i) cannot hold for more than three triples (i, j, k) with $i + j + k = 0$. Indeed, as observed previously, the intersection points $P_{ijk} := p_i^* \cap q_j^* \cap r_k^*$ are all distinct, and so the lines containing $\{p_i, q_j, r_k\}$ are distinct for distinct triples (i, j, k) .

Let Ω be the set of triples (i, j, k) with $i + j + k = 0, -L \leq i \leq L, -2L \leq j < -L$ and $L < k \leq 2L$. Suppose that (i) holds for some triple $(i, j, k) \in \Omega$ and that p_i, q_j, r_k all lie on ℓ (say). We now consider the triples $(i', j', k') \in \Omega$ with $i \neq i', j \neq j', k \neq k'$. With at most two exceptions, (ii) holds for any such triple. Fix one of these triples for which (ii) holds. One of the points $p_{i'}, q_{j'}, r_{k'}$ then lies on ℓ . Suppose that $p_{i'}$ lies in ℓ . But, noting that $1 \leq -i' - j \leq m$, we see that $p_{i'}, q_j$ and $r_{-i'-j}$ are collinear and so $r_{-i'-j}$ lies on ℓ as well. That is, the lines containing $\{p_i, q_j, r_k\}$ and $\{p_{i'}, q_j, r_{-i'-j}\}$ are the same. This is a contradiction as we noted above. Similarly, if $q_{j'}$ lies in ℓ , then $1 \leq -i - j' \leq m$ and we can conclude that the lines containing $\{p_i, q_j, r_k\}$ and $\{p_i, q_{j'}, r_{-i-j'}\}$ are again coincident, a contradiction. Finally, if $r_{k'}$ lies in ℓ , then $-m \leq -i - k' \leq 1$ and the lines containing $\{p_i, q_j, r_k\}$ and $\{p_i, q_{-i-k'}, r_{k'}\}$ are coincident, again a contradiction.

It follows that, whenever $(i, j, k) \in \Omega$, we are in case (ii) and not in (i), that is to say the points p_i, q_j, r_k lie on one of each of the lines ℓ, ℓ', ℓ'' , but do not all lie on one of the lines $\ell, \ell',$ or ℓ'' . Suppose without loss of generality that $p_0 \in \ell, q_{-2L} \in \ell', r_{2L} \in \ell''$. If $q_{-2L+1} \in \ell''$ then the concurrent lines $p_{-1}, q_{-2L+1}, r_{2L}$ all lie in ℓ'' ; as $(-1, -2L + 1, 2L) \in \Omega$, we obtain a contradiction. Similarly, if $q_{-2L+1} \in \ell$, then p_0, q_{-2L+1}, r_{2L+1} all lie in ℓ , again a contradiction. Thus $q_{-2L+1} \in \ell'$, which implies $r_{2L-1} \in \ell''$. Repeating this argument we see that in fact all of the points $q_j, -2L \leq j < -L$, lie on ℓ' and all of the points $r_k, L < k \leq 2L$, lie on ℓ'' . Finally, considering the triple $(i, -2L - i, 2L) \in \Omega$, we see that all of the points $p_i, 0 \leq i < L$, lie on ℓ . We have established that each of the lines ℓ, ℓ', ℓ'' contains at least L of the points p_i, q_j, r_k , concluding the proof of the lemma in Case 1.

We turn now to Case 2, where the argument is very similar. Consider once again a triple of points p_i, q_j, r_k with $i + j + k = 0$. These lie on a line. By Bézout’s theorem, there are two cases:

- (i) p_i, q_j, r_k all lie on ℓ ;
- (ii) two of p_i, q_j, r_k lie on σ and the other lies on ℓ .

If (i) ever holds for some triple (i, j, k) then there is at most one such triple. Suppose it holds for some triple $(i, j, k) \in \Omega$. There are again many triples $(i', j', k') \in \Omega$ with $i \neq i', j \neq j', k \neq k'$. For any such triple, two of $p_{i'}, q_{j'}, r_{k'}$ lie on σ and the other lies on ℓ . Suppose, without loss of generality, that $p_{i'} \in \ell$. Then, noting that

$1 \leq -i' - j \leq m$, we see that $p_{i'}, q_j, r_{-i'-j}$ are collinear and so $r_{-i'-j}$ lies on ℓ as well, and thus (i) also holds for the triple $(i', j, -i' - j)$. This leads to a contradiction exactly as before.

It follows that, whenever $(i, j, k) \in \Omega$, two of the points p_i, q_j, r_k lie on σ and the other lies on ℓ . If $p_0 \in \sigma$ then one of q_{-j}, r_j lies on σ and the other lies on ℓ , for each j with $L < j \leq 2L$. Thus both σ and ℓ contain at least L of the points p_i, q_j, r_k . If, on the other hand, $p_0 \in \ell$, then all the q_{-j}, r_j with $L < j \leq 2L$ lie on σ . But for each $i, |i| \leq L - 1$, there are some j, k with $-2L \leq j < -L$ and $L < k \leq 2L$ such that $(i, j, k) \in \Omega$, and so $p_i \in \ell$ for all these i too. Thus both ℓ and σ contain at least L of the points p_i, q_j, r_k in this case also. □

We remark that this analysis can be pushed further in order to say something about the distribution of the points p_i, q_j, r_k on (for example) three lines ℓ, ℓ', ℓ'' . One could most probably give some kind of complete classification of (say) $100 \times 100 \times 100$ triangular grids. However it is also possible to take a self-contained additive-combinatorial approach, leading to better bounds, and this is the technique we pursue in Sect. 6.

To conclude this section let us remark that a number of beautiful pictures of triangular structures arising from cubic curves (for various different types of cubic) may be found in the paper [6], another work in the interpolation theory literature.

5 Almost Triangular Structure and Covering by Cubics

Recall that if $P \subset \mathbb{RP}^2$ is a set of points then Γ_P is the graph defined by the dual lines $p^*, p \in P$. We now know (Proposition 3.1) that if P has few ordinary lines then Γ_P has a highly triangular structure. We also understand (Lemma 4.4) that triangular structure in Γ_P corresponds to points of P lying on a cubic curve. In this section, we put these facts together to prove some of the structural results stated in the introduction.

The main result is Lemma 5.2, whose statement and proof are somewhat technical. To convey the main idea (and because we will need it later, and because it may be of independent interest) we first establish the following much easier result. This result also comes with better bounds—indeed it says something even if one only knows that P spans $o(n^2)$ ordinary lines—than our more technical later result.

Proposition 5.1 (Cheap structure theorem) *Suppose that P is a finite set of n points in the plane. Suppose that P spans at most Kn ordinary lines for some $K \geq 1$. Then P lies on the union of $500K$ cubic curves.*

Proof We first dispose of a degenerate case. Suppose that one of the dual lines $p^*, p \in P$, meets fewer than $500K$ points in Γ_P . Then every dual line meets one of these points, which means that P is covered by at most $500K$ lines. As every line is already a cubic curve, we are done in this case. Thus we may assume that each dual line p^* meets at least $500K$ points in Γ_P . In particular, it meets at least three points of Γ_P .

Recall the definition of a “good edge” of Γ_P : an edge both of whose vertices have degree 6, and where both faces adjoining it are triangles.

Let us say that an edge is *really good* if all paths of length two from both of its endpoints consist entirely of good edges. If we have a segment S of $l \geq 1$ consecutive

edges on p^* , all of which are really good, then the structure of Γ_P is locally that of a triangular grid with dimensions $\{-2, \dots, 2\}, \{-l - 4, \dots, -1\}, \{1, \dots, l + 4\}$; note that the distinct intersection property of Definition 4.2(i) is automatic since every dual line is assumed to meet at least three points in Γ_P . Applying Lemma 4.4, we conclude that if S is such a segment of consecutive really good edges, containing at least one edge, then the set of $q \in P \setminus \{p\}$ for which q^* meets S all lie on a cubic curve γ_S (which also contains p).

If an edge is not really good, we say that it is *somewhat bad*. We know, by Proposition 3.1, that the number of bad edges is at most $16Kn$. Now associated to any somewhat bad edge e is a path of length 1, 2 or 3 whose first edge is e and whose last edge is bad, and which is furthermore the only bad edge on that path (take a minimal path starting in e and ending in a bad edge). The number of paths of length 3 of the form bad–good–good is at most $16Kn \times 5 \times 5$, since each vertex of a good edge has degree 6. Taking account of paths of length 2 and 1 as well, we obtain an upper bound of $500Kn$ for the number of somewhat bad edges.

By the pigeonhole principle there is a line p^* which contains $t \leq 500K$ somewhat bad edges. These somewhat bad edges partition p^* into t segments of consecutive really good edges (a segment may have length zero). Let the segments with at least one edge be $S_1, \dots, S_{t'}$, and let the segments of length zero, which are simply vertices, consist of vertices $v_{t'+1}, \dots, v_t$.

If $q \in P \setminus \{p\}$, then q^* meets p^* either in a vertex of one of the S_i , or in one of the additional vertices v_j . In the former case, as discussed previously, Lemma 4.4 places q in a cubic curve γ_{S_i} depending on S_i . In the latter case, q lies in the dual line v_j^* . Such a dual line can be thought of as a (degenerate) cubic curve. Taking the union of all these cubic curves, of which there are at most $t' + (t - t') \leq 500K$, gives the result. □

Proposition 5.1 is already a fairly strong structure theorem for sets with few ordinary lines. It is possible that the ordinary lines in a union of $O(1)$ cubics can be analysed directly, though this certainly does not seem to be straightforward. Fortunately, there is much more to be extracted from Proposition 3.1 and the results of Sect. 4, enabling us to prove more precise statements that refine Proposition 5.1, albeit with somewhat worse explicit constants.

The next lemma is the main technical result of this section. In an effort to make the paper more readable, we have formulated it so that, once it is proven, we will have no further need of the dual graph Γ_P and consequences of Melchior’s inequality.

Lemma 5.2 *Suppose that P is a set of n points in the plane. Suppose that P spans at most Kn ordinary lines for some $K \geq 1$, and let $L \geq 10$ be a parameter. Suppose that P cannot be covered by a collection of $4L$ concurrent lines. Then for every $p \in P$ there is a partition $P = \{p\} \cup \Sigma_{1,p} \cup \dots \cup \Sigma_{c_p,p}$ with the following properties:*

- (i) *For $i = 1, \dots, c_p$ the points of $\Sigma_{i,p}$ lie on a (not necessarily irreducible) curve $\gamma_{i,p}$ of degree at most three, which also contains p ;*
- (ii) *If $\gamma_{i,p}$ is not a line, then each irreducible component of it contains at least L points of P ;*

- (iii) If $\gamma_{i,p}$ is not a line, then the points of $\Sigma_{i,p}$ may be partitioned into pairs (q, r) such that p, q, r are collinear, and no other points of P are on the line joining p, q and r ;
- (iv) We have the upper bound $\sum_{p \in P} c_p \leq 2^{19} L^3 K n$ on the average size of c_p .

Proof The proof of this lemma is basically the same as the last one, except now we work with a considerably enhanced notion of what it means to be a “really good edge”. Call an edge *extremely good* if all paths of length $2L$ from both of its endpoints consist entirely of good edges. In the last proposition, we only needed paths of length 2. If an edge e is not extremely good, let us say that it is *slightly bad*.

We now count the number of slightly bad edges by an argument similar to that used to prove the previous proposition. Let e be a slightly bad edge, and let r be the length of the shortest path from an endpoint of e to a vertex of a bad edge. Thus $0 \leq r \leq 2L - 1$, and there is a vertex v of a bad edge e' that is at distance exactly r from an endpoint w of e . Then all paths of length up to r from either of the vertices of e are good, which means that the r -neighbourhood of e has the combinatorial structure of a triangular grid, and also that v lies on the boundary of this neighbourhood and has degree six. Among other things, this implies that among all the paths of length r from v to w , there is a path that changes direction only once. To describe this path, as well as the slightly bad edge e , one could specify the bad edge e' , followed by an endpoint v of that bad edge of degree six, followed by an edge emanating from v , which is followed along for some length r_1 to a vertex of degree six, at which point one switches to one of the other four available directions and follows that direction for a further length r_2 , with $r_1 + r_2 \leq r$ (so in particular $0 \leq r_1, r_2 \leq 2L - 1$), until one reaches a vertex w , at which point the slightly bad edge e is one of the six edges adjacent to w . From Proposition 3.1 and simple counting arguments, we may thus bound the total number of slightly bad edges e crudely by

$$16Kn \times 2 \times 6 \times 2L \times 4 \times 2L \times 6 \leq 2^{15} KL^2 n$$

and so we conclude that the number of slightly bad edges is at most $2^{15} KL^2 n$. One could save a few powers of two here by being more careful, but we will not do so.

Suppose that there are b_p slightly bad edges on p^* . Then

$$\sum_{p \in P} b_p \leq 2^{15} L^2 K n. \tag{5.1}$$

If we have a segment of $m \geq 10L$ consecutive edges on p^* , all of which are extremely good, and with p^* containing at least $4L$ additional edges beyond these m , then the structure of Γ_P is locally that of a triangular grid of dimensions $\{-2L, \dots, 2L\}, \{-m, \dots, -1\}, \{1, \dots, m\}$. Note that as we are assuming that P cannot be covered by $4L$ concurrent lines, every dual line p^* meets at least $4L + 1$ distinct points in Γ_P , ensuring the disjointness property in Definition 4.2(ii). Indeed, the fact that each dual line meets at least $4L + 1$ distinct points, and that p^* contains at least $4L$ additional edges beyond the m consecutive edges, ensures that all the lines in this triangular grid are distinct.

Thus if S is such a segment then, by Lemma 4.5, the set Σ_S of all $q \in P \setminus \{p\}$ for which q^* meets S all lie on a cubic curve γ_S which contains p , and each component of which contains at least L points of P . Furthermore, since the lines q^* meet the vertices of S in pairs (since each such vertex certainly has degree 6) the points of Σ_S may be divided into pairs (q, r) such that p, q, r are collinear, and no other point of P lies on the line joining p, q and r . Compare with conclusion (iii) of this lemma.

The line p^* is divided into b_p segments, each containing one or more vertices, by the slightly bad edges. We then create some subsegments S_1, \dots, S_t by the following rule:

- (i) If p^* contains at most $14L$ edges in all, then we set $t = 0$, so no subsegments S_1, \dots, S_t are created;
- (ii) If p^* contains more than $14L$ edges, and one of the segments S cut out by the slightly bad edges contains all but at most $4L$ of the edges, we set $t = 1$, and define S_1 to be a subsegment of S omitting precisely $4L$ edges;
- (iii) In all other cases, we set S_1, \dots, S_t be those segments cut out by the slightly bad edges with at least $10L$ edges.

We then let v_{t+1}, \dots, v_{c_p} be the vertices not contained in any of the S_1, \dots, S_t . By construction, we see that we always have $t \leq b_p$ and that the number $c_p - t$ of remaining vertices v_i is at most $\max(14L, 4L, (10L + 1)b_p) \leq (10L + 1)b_p + 14L$. We thus have

$$c_p \leq (10L + 2)b_p + 14L. \tag{5.2}$$

Define $\Sigma_{i,p} := \Sigma_{S_i}$ and $\gamma_{i,p} := \gamma_{S_i}$ for $i \leq t$, and for $i \geq t + 1$ let $\gamma_{i,p}$ be the line v_i^* and take $\Sigma_{i,p}$ to consist of the points of $P \setminus \{p\}$ lying on this line.

This collection of cubics and lines has properties (i), (ii) and (iii) claimed in the lemma. The bound (iv) follows immediately from (5.1) and (5.2) and the crude bound $(10L + 2)2^{15}L^2K + 14L \leq 2^{19}L^3K$, valid for $L \geq 10$ and $K \geq 1$. □

We are now in a position to prove a result which is still not quite as strong as our main structure theorem, Theorem 1.5, but is still considerably more powerful (albeit with worse explicit constants) than the rather crude statement of Proposition 5.1.

Proposition 5.3 (Intermediate structure theorem) *Suppose that P is a finite set of n points in the plane. Suppose that P spans at most Kn ordinary lines for some $K \geq 1$. Then one of the following three alternatives holds:*

- (i) P is contained in the union of an irreducible cubic γ and an additional $2^{75}K^5$ points.
- (ii) P lies on the union of an irreducible conic σ and an additional $2^{64}K^4$ lines. Furthermore, σ contains between $\frac{n}{2} - 2^{76}K^5$ and $\frac{n}{2} + 2^{76}K^5$ points of P , and $P \setminus \sigma$ spans at most $2^{62}K^4n$ ordinary lines.
- (iii) P is contained in the union of $2^{16}K$ lines and an additional $2^{87}K^6$ points.

Remark The explicit expressions such as $2^{75}K^5$ in the above proposition could of course be replaced by the less specific notation $O(K^{O(1)})$ if desired, and the reader may wish to do so in the proof below as well.

Proof If P can be covered by $60000K \leq 2^{16}K$ concurrent lines then we are of course done, so we will assume that this is not the case.

By Proposition 5.1 we know that P is covered by at most $500K$ cubic curves. By breaking each of these curves up into irreducible components, we may thus cover P by distinct irreducible cubic curves $\gamma_1, \dots, \gamma_m$ for some

$$m \leq 1500K. \tag{5.3}$$

By Bézout’s Theorem, no pair of distinct irreducible curves intersects in more than 9 points, and so there is a set $P' \subset P$, with

$$|P \setminus P'| \leq 9 \binom{m}{2} \leq 2^{24}K^2,$$

such that each point of P' lies on just one of the curves γ_i .

Suppose first of all that one of the γ_i , say γ_1 , is an irreducible cubic and contains at least $2^{76}K^5$ points of P . Then it also contains at least $2^{75}K^5$ points of P' . Write $n_0 := |P' \cap \gamma_1|$; thus $n_0 \geq 2^{75}K^5$.

By construction and (5.3), P is not covered by $40m$ concurrent lines. Applying Lemma 5.2 with $L := 10m$, we see that for each $p' \in P$ we may partition P as $\{p'\} \cup \Sigma_{1,p'} \cup \dots \cup \Sigma_{c_{p'},p'}$, where $\sum_{p' \in P} c_{p'} \leq 2^{19}(10m)^3Kn \leq 2^{61}K^4n$ and each $\Sigma_{i,p'}$ is contained in some (not necessarily irreducible) cubic $\gamma_{i,p'}$ containing p' which is either a line, or has the property that each irreducible component of it contains at least $10m$ points of P .

By the pigeonhole principle, there is some $p' \in P' \cap \gamma_1$ with the property that $c_{p'} \leq 2^{61}K^4n/n_0$. Fix this p' . By Bézout’s theorem, an irreducible curve of degree at most three that is not already one of the γ_j meets P in no more than $9m$ points, and so we infer that each $\gamma_{i,p'}$ is either a line, or else every irreducible component of it is one of the γ_j . Since p' lies on γ_1 but not on any other γ_i , we infer that all the $\gamma_{i,p'}$ are lines except that one of them, say $\gamma_{1,p'}$, may be γ_1 . Furthermore, none of the lines $\gamma_{j,p'}$, $j = 2, \dots, c_{p'}$, which all contain p' , coincides with any of $\gamma_2, \dots, \gamma_m$. By another application of Bézout’s theorem, each of them contains at most $3m$ points of P .

It follows that

$$\begin{aligned} n &= |P| \leq |P \cap \gamma_1| + \sum_{j=2}^{c_{p'}} |P \cap \gamma_{j,p'}| \\ &\leq n_0 + 2^{24}K^2 + 3mc_{p'} \leq n_0 + \frac{2^{74}K^5n}{n_0}. \end{aligned}$$

Since $n_0 \geq 2^{75}K^5$, we conclude that $n_0 \geq n/2$, which when inserted again into the above inequality gives $n_0 \geq n - 2^{75}K^5$, which is option (i) of Proposition 5.3.

The analysis of option (ii) goes along similar lines but is a little more complicated. Suppose now that one of the γ_i , say γ_1 , is an irreducible conic and contains at least $2^{76}K^5$ points of P . Once again, it also contains at least $2^{75}K^5$ points of P' . Write $n_0 := |P' \cap \gamma_1|$; thus $n_0 \geq 2^{75}K^5$.

By Lemma 5.2 as before we may, for each $p' \in P$, partition P as $\{p'\} \cup \Sigma_{1,p'} \cup \dots \cup \Sigma_{c_{p'},p'}$ with $\sum_{p'} c_{p'} \leq 2^{19}(10m)^3Kn \leq 2^{61}K^4n$ and each $\Sigma_{i,p'}$ contained in some (not necessarily irreducible) curve $\gamma_{i,p'}$ of degree at most three containing p' which is either a line, or has the property that each irreducible component contains at least $10m$ points of P .

By the pigeonhole principle as before, we may find $p' \in P' \cap \gamma_1$ such that $c_{p'} \leq 2^{61}K^4n/n_0$. Now fix this p' .

Suppose that $\gamma_{i,p'}$ is not a line. Then, by Bézout’s theorem as above, each irreducible component of $\gamma_{i,p'}$ is one of the γ_j . Since $p' \in \gamma_{i,p'}$, and p' lies on γ_1 but not on any other γ_j , one of the irreducible components of $\gamma_{i,p'}$ is γ_1 . Thus for each i one of the following is true:

- (i) $\Sigma_{i,p'}$ is contained in a line through p' ;
- (ii) $\Sigma_{i,p'}$ is contained in the conic γ_1 ;
- (iii) $\Sigma_{i,p'}$ is contained in the union of the conic γ_1 and a line γ_{j_i} .

Now recall Lemma 5.2. Item (iii) of that lemma asserts that in cases (ii) and (iii) above the points of $\Sigma_{i,p'}$ may be divided into collinear triples (p', q, r) . This immediately rules out option (ii). For those i satisfying (iii) we see that $|\Sigma_{i,p'}| = 2|\Sigma_{i,p'} \cap \gamma_1|$. For those i satisfying (i) it follows from Bézout’s theorem that $|\Sigma_{i,p'}| \leq 3m$.

Let I be the set of indices i for which $\Sigma_{i,p'}$ is not contained in a line through p' , that is to say for which option (iii) above holds. It follows that

$$n = |P| = 1 + \sum_{i=1}^{c_{p'}} |\Sigma_{i,p'}| \leq 1 + 2 \sum_{i \in I} |\Sigma_{i,p'} \cap \gamma_1| + 3mc_{p'}$$

However, any line through p' meets γ_1 (which contains p') in at most one other point, and so

$$\sum_{i \in I} |\Sigma_{i,p'} \cap \gamma_1| \leq |P \cap \gamma_1| + c_{p'}$$

Since

$$|P \cap \gamma_1| \leq |P' \cap \gamma_1| + |P \setminus P'| \leq n_0 + 2^{24}K^2$$

we conclude that

$$n \leq 2n_0 + 2^{25}K^2 + (3m + 2)c_{p'} + 1 \leq 2n_0 + \frac{2^{74}K^5n}{n_0}$$

Since $n_0 \geq 2^{75}K^5$, this is easily seen to imply that $n_0 \geq n/4$, and hence $n_0 \geq n/2 - 2^{76}K^5$ and $c_{p'} \leq 2^{63}K^4$. In the converse direction, we have

$$n \geq \sum_{i \in I} |\Sigma_{i,p'}| = 2 \sum_{i \in I} |\Sigma_{i,p'} \cap \gamma_1|$$

and so

$$|P \cap \gamma_1| \leq \frac{n}{2} + 1 + \left| \bigcup_{i \notin I} \Sigma_{i,p'} \cap \gamma_1 \right|.$$

For $i \notin I$, $\Sigma_{i,p'} \cap \gamma_1$ consists of at most one point, and so

$$|P \cap \gamma_1| \leq \frac{n}{2} + 1 + c_{p'} \leq \frac{n}{2} + 2^{76}K^5$$

and hence γ_1 contains between $\frac{n}{2} - 2^{76}K^5$ and $\frac{n}{2} + 2^{76}K^5$ elements of P .

Looking back at the three possibilities (i), (ii) and (iii) above, we see that the other points lie in the union of the lines $\gamma_{i,p'}$ and γ_j , of which there are at most $c_{p'} + m \leq 2^{64}K^4$.

To complete the proof that we are in case (ii) claimed in the proposition, we need to give an upper bound for the number of ordinary lines spanned by the set $P \setminus \gamma_1$. Such a line could be ordinary in P , but there are at most Kn such lines. Otherwise, such a line passes through a point $p' \in P \cap \gamma_1$, and contains precisely two points in $P \setminus \gamma_1$. Let us say that such a line is *bad*. The number of bad lines arising from $p' \in P \setminus P'$, a set of cardinality at most $2^{24}K^2$, is at most $2^{23}K^2n$. Suppose then that $p' \in P' \cap \gamma_1$. As above, for each such p' we have a partition $P = \{p'\} \cup \Sigma_{1,p'} \cup \dots \cup \Sigma_{c_{p'},p'}$, and we now know that $\Sigma_{i,p'}$ is either contained in a line through p' or is contained in the union of γ_1 and a line. Furthermore in the latter case we know from Lemma 5.2 (iii) that every line though p' and a point of $\Sigma_{i,p'}$ passes through precisely two other points of P , one on γ_1 and the other not. Therefore it is not bad. The number of bad lines through p' is thus at most $c_{p'}$, and so the total number of bad lines arising from $p' \in P \cap \gamma_1$ is at most $\sum_{p'} c_{p'} \leq 2^{61}K^4n$. Statement (ii) of the proposition follows immediately.

We have now considered all cases in which any irreducible cubic or conic from amongst the m curves γ_i contains more than $2^{76}K^5$ points of P . If this is not the case, the only curves among the γ_i containing more than $2^{76}K^5$ points of P are lines. Thus P may be covered by $m \leq 2^{16}K$ lines and at most $2^{76}K^5m \leq 2^{87}K^6$ points, which gives option (iii). □

6 Unions of Lines

Suppose that P is a set of n points spanning at most Kn ordinary lines. We know from Proposition 5.3 that all but $O(K^{O(1)})$ points of P lie on an irreducible cubic, an irreducible conic and some lines, or some lines. The aim of this section is to reduce the number of lines, in all cases, to at most one. The main result of this section is the following theorem, which may again be of independent interest.

Proposition 6.1 *Suppose that a set $P \subset \mathbb{RP}^2$ of size n lies on a union $\ell_1 \cup \dots \cup \ell_m$ of lines, and that P spans at most Kn ordinary lines. Suppose that $n \geq n_0(m, K)$ is sufficiently large. Then all except at most $3K$ of the points of P lie on a single line.*

We first handle the (easy) case where one line has almost all the points. For this, one does not need to know that P is contained in the union of a few lines.

Lemma 6.2 *Suppose that $P \subset \mathbb{RP}^2$ is a set of size n , that P spans at most Kn ordinary lines, and that at least $\frac{2}{3}n$ of the points of P lie on a single line ℓ . Then in fact all except at most $3K$ of the points lie on ℓ .*

Proof Let $p \in P \setminus \ell$. Then p forms at least $2n/3$ lines with the points of $P \cap \ell$. At most $n/3$ of these contain another point of P , and so at least $n/3$ of them are ordinary. Therefore the number of ordinary lines is at least $|P \setminus \ell|n/3$, and the claim follows immediately. \square

Suppose now, and for the rest of the section, that $P \subset \ell_1 \cup \dots \cup \ell_m$. The opposite extreme to that considered by the above lemma is when all the lines ℓ_i contain many points of P . The next result, which is the key technical step in the proof of Proposition 6.1 (and is in fact rather stronger than that proposition), may be of independent interest.

Proposition 6.3 *Suppose that $m \geq 2$ and that a set $P \subset \mathbb{RP}^2$ of size n lies on a union $\ell_1 \cup \dots \cup \ell_m$ of lines, and that at least εn points of P lie on each of the lines ℓ_i . Suppose that $m, \frac{1}{\varepsilon} \leq n^{\frac{1}{10000}}$. Then P spans at least $\varepsilon^{12}n^2/m^6$ ordinary lines.*

Remark The exponent $\frac{1}{10000}$ could certainly be improved somewhat, but a really significant improvement—beyond $\frac{1}{100}$, say—would require new methods.

The proof of this proposition is quite long. Before embarking upon it we show how to derive Proposition 6.1 as a consequence.

Deduction of Proposition 6.1 from Proposition 6.3 and Lemma 6.2 Reorder the lines so that $n_1 \geq n_2 \geq \dots \geq n_m$, where $n_i := |P \cap \ell_i|$. Set $\varepsilon_j := n_j/n$. If $\varepsilon_2 \leq 1/3m$ then $\varepsilon_1 \geq 2/3$ and we are done by Lemma 6.2, so suppose that $\varepsilon_2 \geq 1/3m$. Write $P_j := P \cap (\ell_1 \cup \dots \cup \ell_j)$, $j = 2, 3, \dots, m$. By Proposition 6.3, the set P_j determines at least $\varepsilon_j^{12}n^2/m^6$ ordinary lines (here we have used the trivial lower bound $|P_j| \geq |P_1| \geq n/m$). Since $|P \setminus P_j| \leq m\varepsilon_{j+1}n$, the number of these which fail to be ordinary lines in P is bounded above by $m\varepsilon_{j+1}n^2$. Let j be the least index such that $\varepsilon_{j+1} < \frac{1}{2}\varepsilon_j^{12}m^{-7}$ or, if there is no such index, set $j := m$. Then it follows that the number of ordinary lines in P is at least $\frac{1}{2}\varepsilon_j^6n^2/m^5$. We have the lower bound $\varepsilon_j \geq \exp(-e^{Cm})$, and so P spans $\gg \exp(-e^{Cm})n^2$ ordinary lines. If $n \geq n_0(m, K)$ is sufficiently large this is greater than Kn , so we obtain a contradiction. \square

Remark We note that $n_0(m, K)$ can be taken to have the shape $n_0(m, K) \sim K \exp \exp(Cm)$.

We may now focus our attention on establishing Proposition 6.3. We will divide into two quite different cases, according as the lines ℓ_i all intersect at a point or not.

Proposition 6.4 *Suppose that $m \geq 2$ and that P lies on a union $\ell_1 \cup \dots \cup \ell_m$ of lines, all of which pass through a point, and that at least εn points of P lie on each of the lines ℓ_i . Then P spans at least $\varepsilon^2n^2/50$ ordinary lines.*

Proof We are greatly indebted to Luke Alexander Betts, a second year undergraduate at Trinity College, Cambridge, who showed us the following argument. For brevity we give a slightly crude version of the argument he showed us.

Applying a projective transformation, we may assume without loss of generality that all the lines pass through the origin $[0, 0, 1]$ in the affine part of \mathbb{RP}^2 . Dualising, these m lines become points on the line at infinity, and the sets $P \cap \ell_i$ become sets of parallel lines. If \mathcal{L} is a set of lines in \mathbb{R}^2 , we say that a point is *ordinary* for \mathcal{L} if it lies on precisely two of the lines in \mathcal{L} . We say that \mathcal{L} is t -parallel if, for every line $\ell \in \mathcal{L}$, there are at least t other lines parallel to it. Finally, we say that a point lying on three or more of the lines from \mathcal{L} is a *triple point*. The dual statement to Proposition 6.4 (with t replacing εn) is then the following. \square

Proposition 6.5 *Let $t > 0$ be a real number. Suppose that \mathcal{L} is a t -parallel set of lines in \mathbb{R}^2 , and that not all the lines of \mathcal{L} are parallel. Then there are at least $t^2/50$ ordinary points for \mathcal{L} .*

The heart of the matter is the following lemma.

Lemma 6.6 *Suppose that \mathcal{L} is t -parallel, but not all the lines of \mathcal{L} are parallel. Then there is a line $\ell \in \mathcal{L}$ containing at least $t/2$ ordinary points for \mathcal{L} .*

Proof If there are no triple points determined by \mathcal{L} then the conclusion is immediate, as every line intersects at least $t + 1 > t/2$ other lines. If there are triple points determined by \mathcal{L} , let T be the set of them. Let v be a vertex of the convex hull of T , lying on lines $\ell_1, \ell_2, \ell_3 \in \mathcal{L}$. There are open rays (half-lines) $\ell_1^+, \ell_2^+, \ell_3^+$ emanating from v which do not intersect T . Suppose without loss of generality that the rays ℓ_1^+, ℓ_3^+ lie on either side of the line ℓ_2 , as depicted in Fig. 15. Each of the t other lines in \mathcal{L} parallel to ℓ_2 meets either ℓ_1^+ or ℓ_3^+ , and so at least one of these rays meets at least $t/2$ lines in \mathcal{L} . All of these points of intersection must, by construction, be ordinary points for \mathcal{L} .

Now let us return to the main problem, the dual form of Proposition 6.4 stated above. Note that if \mathcal{L}' is formed by removing at most $t/5$ lines from \mathcal{L} then it is still $4t/5$ -parallel. Thus, by $\lceil t/5 \rceil$ applications of Lemma 6.6 we may inductively find distinct lines $\ell_1, \dots, \ell_{\lceil t/5 \rceil}$ such that ℓ_i contains at least $2t/5$ ordinary points for $\mathcal{L} \setminus \{\ell_1, \dots, \ell_{i-1}\}$. These are not necessarily ordinary points for \mathcal{L} , but any such point that is not lies on one of $\ell_1, \dots, \ell_{i-1}$. Since it also lies on ℓ_i , there are at most $i - 1 < t/5$ such points, and so ℓ_i contains at least $t/5$ ordinary points of \mathcal{L} . Each ordinary point of \mathcal{L} lies on at most two of the lines ℓ_i , so we get at least $t^2/50$ ordinary points in total. \square

We have now established Proposition 6.4, which is the particular case of Proposition 6.3 in which all the lines ℓ_i pass through a single point. We turn now to the case in which this is not so. The next proposition, together with Proposition 6.4, immediately implies Proposition 6.3 and hence the main result of the section, Proposition 6.1.

Proposition 6.7 *Suppose that $m \geq 2$ and that a set $P \subset \mathbb{RP}^2$ of size n lies on a union $\ell_1 \cup \dots \cup \ell_m$ of lines, not all of which pass through a single point, and that at least εn points of P lie on each of the lines ℓ_i . Suppose that $m, \frac{1}{\varepsilon} \leq n^{\frac{1}{10000}}$. Then P spans $\gg \varepsilon^{12} n^2 / m^6$ ordinary lines.*

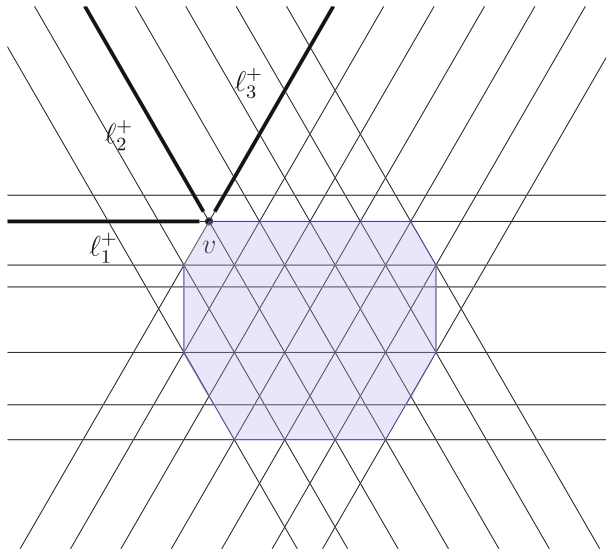


Fig. 15 Figure relevant to the proof of Lemma 6.6. Here, $t = 6$ and the ray ℓ_3^+ meets $4 > 6/2$ lines in \mathcal{L} . All of these points of intersection are ordinary as they lie outside the convex hull of the triple points of \mathcal{L} , shaded in blue (Color figure online)

The proof proceeds via several lemmas. It also requires some additive-combinatorial ingredients not needed elsewhere in the paper, which we collect in the appendix. It is convenient, for this portion of the argument, to work entirely in the affine plane. Let us begin, then, by supposing that a projective transformation has been applied so that all lines ℓ_i and their intersections lie in the affine plane \mathbb{R}^2 .

Suppose that ℓ is a line. Then by a *ratio map* on ℓ we mean a map

$$\psi = \psi_{q,q'} : \ell \rightarrow \mathbb{R} \cup \{\infty\}$$

of the form

$$\psi_{q,q'}(p) = \frac{\text{length}(pq)}{\text{length}(pq')}$$

where q, q' are distinct points on ℓ and the lengths $\text{length}(pq)$, $\text{length}(pq')$ are *signed* lengths on ℓ . We say that the ratio maps $\psi_{q,q'}$ and $\psi_{q',q}$ are *equivalent*, but otherwise all ratio maps are deemed inequivalent. Note that such ratio maps implicitly appeared in the analysis of the infinite near-counterexample (2.2).

An ordered triple of lines ℓ_i, ℓ_j, ℓ_k not intersecting in a single point defines two ratio maps $\phi_{i,j,k} : \ell_i \rightarrow \mathbb{R} \cup \{\infty\}$ and $\tilde{\phi}_{i,j,k} : \ell_j \rightarrow \mathbb{R} \cup \{\infty\}$ via

$$\phi_{i,j,k} := \psi_{\ell_i \cap \ell_j, \ell_i \cap \ell_k} \quad \text{and} \quad \tilde{\phi}_{i,j,k} := \psi_{\ell_i \cap \ell_j, \ell_j \cap \ell_k}.$$

We will make considerable use of these maps in what follows, as well as of the following definition.

Definition 6.8 (*Quotient set*) Suppose that $X \subset \mathbb{R} \cup \{\infty\}$ is a set. Then we write $\mathcal{Q}(X)$ for the set of all quotients x_1/x_2 with $x_1, x_2 \in X$ and $x_1, x_2 \notin \{0, \infty\}$.

The following definition depends on the parameter n , which is the number of points in the set P . Since this is fixed throughout the section, we do not indicate dependence on it explicitly.

Definition 6.9 Let A be a finite subset of some line ℓ , and let ψ be a ratio map on ℓ . Then we say that A is a ψ -grid if A is a union of at most $n^{1/30}$ sets S such that $|\mathcal{Q}(\psi(S))| \leq n^{1+1/10}$. We say that A is a *grid* if it is a ψ -grid for some ratio map ψ on ℓ .

Lemma 6.10 Let ℓ be a line, and suppose that ψ and ψ' are inequivalent ratio maps on ℓ . Suppose that A is a ψ -grid and that A' is a ψ' -grid. Then $|A \cap A'| \ll n^{1-1/25}$.

Proof Without loss of generality we may assume that ℓ is the x -axis parametrised as $\{(t, 0) : t \in \mathbb{R}\}$. By abuse of notation we identify A and A' with subsets of \mathbb{R} . The ratio maps ψ, ψ' are given by $\psi(t) = (t + a)/(t + b), \psi'(t) = (t + a')/(t + b')$ with $a \neq b, a' \neq b'$ and $\{a, b\} \neq \{a', b'\}$.

Suppose that A is a ψ -grid and that A' is a ψ' -grid. Write $A = \bigcup_{i=1}^{n^{1/30}} S_i$ and $A' = \bigcup_{j=1}^{n^{1/30}} S'_j$ where $|\mathcal{Q}(\psi(S_i))|$ and $|\mathcal{Q}(\psi'(S'_j))|$ are both at most $n^{1+1/10}$. It suffices to show that $|S_i \cap S'_j| \ll n^{22/25}$.

Suppose that X is the set of all $x \in S_i \cap S'_j$ for which $\psi(x) > 0$. Since X is contained in both S_i and $S'_j, |\mathcal{Q}(\psi(X))|$ and $|\mathcal{Q}(\psi'(X))|$ are both at most $n^{1+1/10}$. Writing $Y := \{\log \psi(x) : x \in X\}$, we see that

$$|Y - Y|, |f(Y) - f(Y)| \leq 2n^{1+1/10}, \tag{6.1}$$

where

$$\begin{aligned} f(y) &= \log \circ \psi' \circ \psi^{-1} \circ \exp(y) \\ &= \log((b - a')e^y + a' - a) - \log((b - b')e^y + b' - a) \\ &= \log(Ae^y + B) - \log(A'e^y + B'), \end{aligned}$$

say, and $f(Y) := \{f(y) : y \in Y, f(y) \text{ is defined}\}$. Note that we do not have $AA' = 0$ and $BB' = 0$ (there are four cases to consider). We thus compute

$$f''(y) = \frac{e^y(b - a)(b' - a')(AA'e^{2y} - BB')}{(Ae^y + B)^2(A'e^y + B')^2}.$$

This is continuous except when $e^y = -B'/A'$ or $-B/A$, and nonzero except when $e^{2y} = BB'/AA'$. It follows that \mathbb{R} may be split into at most 4 pieces on which f

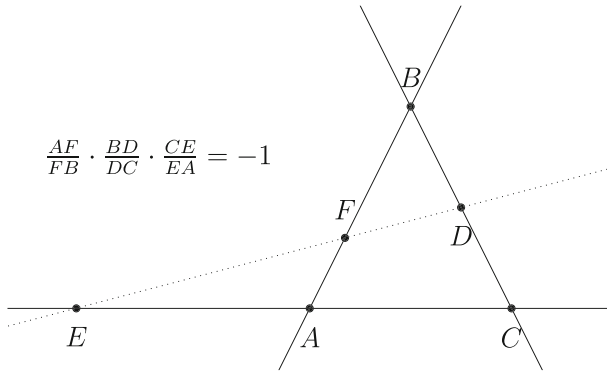


Fig. 16 An illustration of Menelaus’s theorem. The lengths are signed

is defined and strictly concave/convex. By Proposition A.9, this implies that at least one of $|Y - Y|, |f(Y) - f(Y)|$ has size $\gg |Y|^{5/4}$. Comparing with (6.1) we see that $|X| = |Y| \ll n^{2/5}$.

An almost identical argument applies when X is the set of all $x \in S_i \cap S'_j$ for which $\psi(x) < 0$, taking $Y := \{\log(-\psi(x)) : x \in X\}$; now we have $f(y) = \log(-Ae^y + B) - \log(-A'e^y + B')$, but the rest of the argument is the same.

Putting these two cases together gives $|S_i \cap S'_j| \ll n^{2/5}$, which is what we wanted to prove. □

Ratio maps may be used in understanding the metric properties of intersections of lines as a consequence of Menelaus’s theorem, illustrated in Fig. 16.

Lemma 6.11 *Let ℓ_i, ℓ_j, ℓ_k be three lines not meeting at a point. Let $X_i \subset \ell_i$ and $X_j \subset \ell_j$, and let $\Gamma \subset X_i \times X_j$ be a set of pairs, with neither X_i nor X_j containing $\ell_i \cap \ell_j$. Let $X_k \subset \ell_k$ be the set of points on ℓ_k formed by intersecting the lines $\overline{\{x_i, x_j\}}, (x_i, x_j) \in \Gamma$, with ℓ_k . Then*

$$|X_k| = |\{\phi_{i,j,k}(x_i)/\tilde{\phi}_{i,j,k}(x_j) : (x_i, x_j) \in \Gamma\}|.$$

Proof Apply Menelaus’s theorem with $AC = \ell_i, AB = \ell_j$ and $BC = \ell_k$. Suppose that $x_i \in \ell_i$ and $x_j \in \ell_j$, and write $E = x_i, F = x_j$ in the diagram. Then D is the point at which $\overline{\{x_i, x_j\}}$ intersects ℓ_k . Note that $\phi_{i,j,k}(x_i) = EA/EC, \tilde{\phi}_{i,j,k}(x_j) = FA/FB$. By Menelaus’ theorem it follows that $\phi_{i,j,k}(x_i)/\tilde{\phi}_{i,j,k}(x_j) = DB/DC$. This ratio uniquely determines the point D , and the lemma follows. □

Lemma 6.12 *Suppose that P is a set of n points lying on a union $\ell_1 \cup \dots \cup \ell_m$ of lines, that the lines ℓ_i are not all concurrent, and that at least εn points of P lie on each line ℓ_i . Suppose that $m, \frac{1}{\varepsilon} \leq n^{10000}$. Then either P spans at least $\varepsilon^2 n^2/8$ ordinary lines, or else there are at least two values of i such that $P \cap \ell_i$ contains a grid with size $\gg \varepsilon^4 n/m^2$.*

Proof By the dual version of the Sylvester–Gallai theorem, there is some pair of lines ℓ_i, ℓ_j such that no other line passes through $\ell_i \cap \ell_j$. Each of ℓ_i, ℓ_j contains at least

εn points of P , at least $\varepsilon n - 1 \geq \varepsilon n/2$ of which are not the intersection point $\ell_i \cap \ell_j$. If P spans fewer than $\varepsilon^2 n^2/8$ ordinary lines then there is some k such that for at least $\varepsilon^2 n^2/8m$ pairs $p_i \in \ell_i, p_j \in \ell_j (p_i, p_j \neq \ell_i \cap \ell_j)$ the line $\overline{\{p_i, p_j\}}$ meets ℓ_k in a point of P . Write $X_i := (\ell_i \cap P) \setminus (\ell_i \cap \ell_j), X_j := (\ell_j \cap P) \setminus (\ell_i \cap \ell_j)$ and $\Gamma \subset X_i \times X_j$ for the set of pairs $(p_i, p_j) \in X_i \times X_j$ for which $\overline{\{p_i, p_j\}}$ meets ℓ_k in a point of $X_k = \ell_k \cap P$. By Lemma 6.11 it follows that

$$n \geq |X_k| \geq |\{\phi_{i,j,k}(x_i)/\tilde{\phi}_{i,j,k}(x_j) : (x_i, x_j) \in \Gamma\}|.$$

By Corollary A.2 and the hypothesis on m and $\frac{1}{\varepsilon}$ it follows that there are sets $X'_i \subset X_i, X'_j \subset X_j$ with $|X'_i|, |X'_j| \gg \varepsilon^4 n/m^2$ and

$$|\mathcal{Q}(\phi_{i,j,k}(X'_i))|, |\mathcal{Q}(\tilde{\phi}_{i,j,k}(X'_j))| \ll m^{11} n/\varepsilon^{22} < n^{1+\frac{1}{10}}.$$

Thus certainly X'_i is a $\phi_{i,j,k}$ -grid and X'_j is a $\tilde{\phi}_{i,j,k}$ -grid. □

As a result of this lemma we may, in proving Proposition 6.7, restrict attention to sets where $P \cap \ell_i$ contains a large grid for at least two values of i . We study this situation further in the next lemma, whose proof is a little involved.

Lemma 6.13 *Suppose P is a set of n points lying on a union of lines $\ell_1 \cup \dots \cup \ell_m$, where $m \leq n^{\frac{1}{10000}}$. Suppose that $i \neq j$ and that $X_i \subset P \cap \ell_i, X_j \subset P \cap \ell_j$ are grids, both of size at least $\varepsilon' n$, where $\varepsilon' \gg n^{-\frac{1}{1000}}$, and neither containing $\ell_i \cap \ell_j$. Then either P spans $\gg (\varepsilon')^3 n^2$ ordinary lines, or else there is a line ℓ_k , not passing through $\ell_i \cap \ell_j$, and a grid $X_k \subset P \cap \ell_k$, such that all but at most $\varepsilon' |X_i||X_j|/25$ pairs $(x_i, x_j) \in X_i \times X_j$ are such that the line $\overline{\{x_i, x_j\}}$ meets ℓ_k in a point of X_k .*

Proof Write $\eta := \varepsilon'/100$. Suppose that P does not contain at least $(\varepsilon')^3 n^2/100 = \eta(\varepsilon' n)^2$ ordinary lines. Then there are at least $(1 - \eta)|X_i||X_j|$ pairs $(x_i, x_j) \in X_i \times X_j$ such that $\overline{\{x_i, x_j\}}$ meets some other line ℓ_k . Write $\Gamma_k \subset X_i \times X_j$ for the set of pairs $(x_i, x_j) \in X_i \times X_j$ such that $\overline{\{x_i, x_j\}}$ meets ℓ_k in a point of P . Thus $\sum_k |\Gamma_k| \geq (1 - \eta)|X_i||X_j|$. We claim that there is at most one value of k for which $|\Gamma_k| \geq \eta |X_i||X_j|/m$, and that for this k (if it exists) the line ℓ_k does not pass through $\ell_i \cap \ell_j$. Note that if $|\Gamma_k| \geq \eta |X_i||X_j|/m$ then certainly $|\Gamma_k| \geq \delta n^2$, where $\delta = n^{-1/250}$.

This claim is a consequence of the following two facts.

Fact 1 *If ℓ_k passes through $\ell_i \cap \ell_j$ then $|\Gamma_k| < \delta n^2$.*

Fact 2 *If we have two lines $\ell_k, \ell_{k'}$, neither passing through $\ell_i \cap \ell_j$, then at least one of $|\Gamma_k|, |\Gamma_{k'}|$ has size at most δn^2 .*

Proof of Fact 1 Suppose that ℓ_k passes through $\ell_i \cap \ell_j$, but that $|\Gamma_k| \geq \delta n^2$. Here (and in the proof of Fact 2 below) we apply an affine transformation so that ℓ_i is the x -axis and ℓ_j is the y -axis. Suppose that ℓ_k is the line $\{(t, \lambda_k t) : t \in \mathbb{R}\}$. Suppose that $X_i = \{(a, 0) : a \in A\}$ and $X_j = \{(0, b) : b \in B\}$ and, by slight abuse of notation, identify Γ_k with a subset of $A \times B$ in the obvious way. A short computation confirms that the intersection of the line through $(a, 0) \in \ell_i$ and $(0, b) \in \ell_j$ with ℓ_k is

the point $(\frac{1}{1/x+\lambda_k/y}, \frac{\lambda}{1/x+\lambda_k/y})$. By Corollary A.2 there are sets $A' \subset A$ and $B' \subset B$ with $|A'|, |B'| \gg \delta n$ and $|\frac{1}{A'} - \frac{1}{A}| \ll \delta^{-11}n \ll n^{1+\frac{1}{10}}$. Since X_i is a grid, there is some ratio function $\psi(t) = (t + \alpha)/(t + \beta)$, $\alpha \neq \beta$, such that $A' \subset \bigcup_{i=1}^{n^{1/30}} S_i$, where $|\mathcal{Q}(\psi(S_i))| \leq n^{1+\frac{1}{10}}$ for each i . Let A'' be the largest of the intersections $A' \cap S_i$; then

$$|A''| \gg \delta n^{1-\frac{1}{30}}, |\mathcal{Q}(\psi(A''))| \leq n^{1+\frac{1}{10}}, \left| \frac{1}{A''} - \frac{1}{A} \right| \ll n^{1+\frac{1}{10}}.$$

Writing $Y := 1/A''$ and $f(y) := \log(1 + \alpha y) - \log(1 + \beta y)$ we thus have

$$|Y| \gg \delta n^{1-\frac{1}{30}}, |Y - Y|, |f(Y) - f(Y)| \ll n^{1+\frac{1}{10}}. \tag{6.2}$$

However

$$f''(y) = \frac{\beta^2}{(1 + \beta y)^2} - \frac{\alpha^2}{(1 + \alpha y)^2} = (\alpha - \beta) \frac{2\alpha\beta y - (\alpha + \beta)}{(1 + \alpha y)^2(1 + \beta y)^2}$$

is continuous away from $y = -1/\alpha$ and $y = -1/\beta$ and has just one real zero, and therefore one may divide \mathbb{R} into at most 4 intervals on the interior of which f is defined and strictly convex/concave. By Proposition A.9 it follows that one of $|Y - Y|, |f(Y) - f(Y)|$ has size $\gg |Y|^{5/4}$. This comfortably contradicts (6.2) for large n . □

Proof of Fact 2 Suppose that $|\Gamma_k|, |\Gamma_{k'}| \geq \delta n^2$. By Lemma 6.11 and the fact that neither ℓ_k nor $\ell_{k'}$ contains more than n points we have

$$|\{\phi_{i,j,k}(x_i)/\tilde{\phi}_{i,j,k}(x_j) : (x_i, x_j) \in \Gamma_k\}| \leq n$$

and

$$|\{\phi_{i,j,k'}(x_i)/\tilde{\phi}_{i,j,k'}(x_j) : (x_i, x_j) \in \Gamma_{k'}\}| \leq n.$$

Applying Corollary A.2 exactly as before we deduce that there are sets $X_i^{(k)}, X_i^{(k')} \subset X_i$ and sets $X_j^{(k)}, X_j^{(k')} \subset X_j$, all of size $\gg \delta n > n^{1-\frac{1}{25}}$, such that all of $|\mathcal{Q}(\phi_{i,j,k}(X_i^{(k)}))|, |\mathcal{Q}(\phi_{i,j,k'}(X_i^{(k')}))|, |\mathcal{Q}(\tilde{\phi}_{i,j,k}(X_j^{(k)}))|$ and $|\mathcal{Q}(\tilde{\phi}_{i,j,k'}(X_j^{(k')}))|$ have size $\ll \delta^{-12}n < n^{1+\frac{1}{10}}$. Thus $X_i^{(k)}$ is a $\phi_{i,j,k}$ -grid, $X_i^{(k')}$ is a $\phi_{i,j,k'}$ -grid, $X_j^{(k)}$ is a $\tilde{\phi}_{i,j,k}$ -grid and $X_j^{(k')}$ is a $\tilde{\phi}_{i,j,k'}$ -grid. Since X_i and X_j are themselves grids, and since grids corresponding to inequivalent ratio functions intersect in a set of size no more than $n^{1-\frac{1}{25}}$ by Lemma 6.10, we see that $\phi_{i,j,k} \sim \phi_{i,j,k'}$ and $\tilde{\phi}_{i,j,k} \sim \tilde{\phi}_{i,j,k'}$.

It follows immediately from the definition of $\phi_{i,j,k}, \phi_{i,j,k'}, \tilde{\phi}_{i,j,k}, \tilde{\phi}_{i,j,k'}$ that $\ell_i \cap \ell_k = \ell_i \cap \ell_{k'}$ and $\ell_j \cap \ell_k = \ell_j \cap \ell_{k'}$, and therefore $\ell_k = \ell_{k'}$. This is contrary to assumption, and so we have established Fact 2.

This completes the proof of the claim. It follows that there is some k such that at least $(1 - 2\eta)|X_i||X_j|$ pairs $x_i \in X_i, x_j \in X_j$ are such that $\overline{\{x_i, x_j\}}$ meets $P \cap \ell_k$, and furthermore ℓ_k does not pass through $\ell_i \cap \ell_j$. If $x_k \in P \cap \ell_k$, we say that x_k is *well-covered* if there are at least $\eta(\varepsilon')^2 n$ pairs $(x_i, x_j) \in X_i \times X_j$ such that $\overline{\{x_i, x_j\}}$ passes through x_k . Since $|P \cap \ell_k| \leq n$, the number of pairs (x_i, x_j) used up by poorly-covered x_k is no more than $\eta(\varepsilon')^2 n^2 \leq \eta|X_i||X_j|$. Thus, for at least $(1 - 3\eta)|X_i||X_j|$ pairs (x_i, x_j) , the line $\overline{\{x_i, x_j\}}$ meets ℓ_k in a well-covered point x_k . Write $\tilde{X}_k \subset P \cap \ell_k$ for the set of well-covered points; we are going to show that there is a grid X_k occupying almost all of \tilde{X}_k .

Note that, by definition of well-covered, for any $Y \subset \tilde{X}_k$ there is a graph $\Gamma \subset Y \times X_i, |\Gamma| \geq \eta(\varepsilon')^2 |Y||X_i|$, such that each line $\overline{\{y, x_i\}}$ meets X_j whenever $(y, x_i) \in \Gamma$. By Lemma 6.11 and the trivial bound $|X_j| \leq n$, this implies that

$$|\{\phi_{k,i,j}(y)/\tilde{\phi}_{k,i,j}(x_i) : (y, x_i) \in \Gamma\}| \leq n.$$

Suppose that $|Y| \geq \eta n$. Then $|\Gamma| \gg \eta^2(\varepsilon')^3 n^2 \gg n^{2-\frac{1}{200}}$. By Corollary A.2 there are sets $Y' \subset Y$ and $X'_i \subset X_i, |Y'|, |X'_i| \gg n^{1-\frac{1}{200}}$, such that $|\mathcal{Q}(\phi_{i,j,k}(Y'))| < n^{1+\frac{1}{10}}$. Applying this argument repeatedly, we see that all of \tilde{X}_k except for a set of size at most ηn can be covered by disjoint sets Y' with these properties. There are at most $Cn^{\frac{1}{200}} < n^{\frac{1}{30}}$ of these sets, and so the union of them, X_k say, is a $\phi_{i,j,k}$ -grid.

Finally, note that the number of pairs (x_i, x_j) for which $\overline{\{x_i, x_j\}}$ passes through $\tilde{X}_k \setminus X_k$ is at most $n|\tilde{X}_k \setminus X_k| \leq \eta n$. For all other pairs, $\overline{\{x_i, x_j\}}$ passes through the grid X_k .

At last, this completes the proof of the lemma. □

We are now in a position to complete the proof of Proposition 6.7 and hence of all the other results in this section.

Proof of Proposition 6.7 Suppose that P lies on $\ell_1 \cup \dots \cup \ell_m$, with at least εn points on each line and not all of the lines through a single point. Suppose that $m, \frac{1}{\varepsilon} \leq n^{\frac{1}{10000}}$. By Lemma 6.12 we are done unless there are two values $i, j \in \{1, \dots, m\}$ such that $P \cap \ell_i, P \cap \ell_j$ both contain grids of size at least $\varepsilon' n$, with $\varepsilon' \gg \varepsilon^4/m^2 > n^{-\frac{1}{1000}}$. Suppose without loss of generality that ℓ_i contains the largest grid amongst all grids in P ; call this X_i . Let X_j be a grid in $P \cap \ell_j$ of size at least $\varepsilon' n$, and apply Lemma 6.13 to these two grids X_i, X_j . If the first conclusion of that lemma holds then P contains $\gg (\varepsilon')^3 n^2 \gg \varepsilon^{12} n^2/m^6$ ordinary lines, and we are done. Otherwise there is a line ℓ_k , not passing through $\ell_i \cap \ell_j$, and a grid $X_k \subset P \cap \ell_k$, such that for all (x_i, x_j) in a set Γ of size at least $(1 - 4\eta)|X_i||X_j|$, the line $\overline{\{x_i, x_j\}}$ meets ℓ_k in a point of X_k .

By Lemma 6.11 we have

$$|\{\phi_{i,j,k}(x_i)/\tilde{\phi}_{i,j,k}(x_j) : (x_i, x_j) \in \Gamma\}| \leq |X_k|.$$

It follows from Lemma A.4 that

$$|X_k| \geq |X_i| + |X_j| - 4 - 4\sqrt{2\eta|X_i||X_j|}.$$

Since $\eta = \varepsilon'/100$ and n is sufficiently large, this is strictly greater than $|X_i|$, contrary to the assumption that X_i was a grid in P of largest size. \square

In this rather long section, we have proved results about sets P , contained in a union $\ell_1 \cup \dots \cup \ell_m$ of lines, spanning few ordinary lines. They are plausibly of independent interest. Let us now, however, return to our main task and combine what we have established with the main results of previous sections. By combining Proposition 6.1 (using the bound on $n_0(m, K)$ noted after Proposition 6.3) with Proposition 5.3 we obtain the following structural result. From the qualitative point of view at least, this supersedes all previous results in the paper (and, in particular, implies Theorem 1.4 as a corollary).

Proposition 6.14 *Suppose that P is a set of n points in \mathbb{RP}^2 for some $n \geq 100$ spanning at most Kn ordinary lines, where $1 \leq K \leq c(\log \log n)^c$ for some sufficiently small absolute constant c . Then P differs in at most $O(K^{O(1)})$ points from a subset of a set of one of the following three types:*

- (i) *An irreducible cubic curve;*
- (ii) *The union of an irreducible conic and a line;*
- (iii) *A line.*

Proof We apply Proposition 5.3. In case (i) we are already done. In cases (ii) and (iii) we see that, after removing an irreducible conic if necessary, that we have $\geq n/2 - O(K^{O(1)})$ points on $O(K^{O(1)})$ lines determining at most $O(K^{O(1)})n$ ordinary lines. By Proposition 6.1 and the hypothesis $K \leq (\log \log n)^c$, all but $O(K^{O(1)})$ of these points lie on a single line, and the claim follows. \square

7 The Detailed Structure Theorem

We turn now to the proof of the detailed structure theorem, Theorem 1.5. We first establish a slightly weaker version in which the linear bounds $O(K)$ have been relaxed to polynomial bounds $O(K^{O(1)})$. This somewhat weaker statement is already sufficient for our main application, the proof of the Dirac–Motzkin conjecture for large n . At the end of this section we indicate how to recover the full strength of Theorem 1.5.

Theorem 7.1 *Suppose that P is a finite set of n points in the projective plane \mathbb{RP}^2 . Suppose that P spans at most Kn ordinary lines for some $K \geq 1$, and suppose that $n \geq \exp \exp(CK^C)$ for some sufficiently large absolute constant C . Then, after applying a projective transformation if necessary, P differs by at most $O(K^{O(1)})$ points from an example of one of the following three types:*

- (i) *$n - O(K^{O(1)})$ points on a line;*
- (ii) *The set X_{2m} defined in (1.1) for some $m = \frac{1}{2}n - O(K^{O(1)})$;*
- (iii) *A coset $H \oplus x$, $3x \in H$, of some finite subgroup H of the real points on an irreducible cubic curve with H having cardinality $n + O(K^{O(1)})$.*

We now prove this theorem. We already know from Proposition 6.14 that a set with at most Kn ordinary lines must mostly lie on a line, the union of a conic and a line, or

an irreducible cubic curve, and the proof of Theorem 1.5 proceeds by analysing the second two possibilities further.

The key to doing this is the fact that collinearities on (possibly reducible) cubic curves are related to group structure. This is particularly clear in the case of an irreducible cubic, as we briefly discussed in Sect. 2. However, one can see some group structure even in somewhat degenerate cases.

An important ingredient in our analysis is the following result of a fairly standard type from additive combinatorics, which may be thought of as a kind of structure theorem for triples of sets A, B, C with few “arithmetic ordinary lines”.

Proposition A.5 *Suppose that A, B, C are three subsets of some abelian group G , all of cardinality within K of n , where $K \leq \varepsilon n$ for some absolute constant $\varepsilon > 0$. Suppose that there are at most Kn pairs $(a, b) \in A \times B$ for which $a + b \notin C$. Then there is a subgroup $H \leq G$ and cosets $x + H, y + H$ such that $|A\Delta(x + H)|, |B\Delta(y + H)|, |C\Delta(x + y + H)| \leq 7K$.*

This result will not be at all surprising to the those initiated in additive combinatorics, but we do not know of a convenient reference for it. We supply a complete proof in Appendix A.

Suppose that P is mostly contained in a cubic curve γ which is not a line. We subdivide into two cases according to whether γ is irreducible or not.

Lemma 7.2 (Configurations mostly on an irreducible cubic) *Suppose that P is a set of n points in \mathbb{RP}^2 spanning at most Kn ordinary lines. Suppose that all but K of the points of P lie on an irreducible cubic curve γ , and suppose that $n \geq CK$ for a suitably large absolute constant C . Then there is a coset $H \oplus x$ of γ with $3x = x \oplus x \oplus x \in H$ such that $|P\Delta(H \oplus x)| = O(K)$. In particular, γ is either an elliptic curve or an acnodal cubic.*

Proof Let γ^* be the smooth points of γ , which we give a group law as in Sect. 2.

Set $P' := P \cap \gamma^*$. Then $|P'| = |P| + O(K)$, and P' spans at most $O(Kn)$ ordinary lines. If $p_1, p_2 \in \gamma^*$ are distinct then the line joining p_1 and p_2 meets γ^* again in the unique point $\ominus p_1 \ominus p_2$. This assumption implies that $\ominus p_1 \ominus p_2 \in P'$ for all but at most $O(Kn)$ pairs $p_1, p_2 \in P'$. Applying Proposition A.5, it follows that there is a coset $H \oplus x$ of γ^* such that $|P\Delta(H \oplus x)| = O(K)$ and also $|P\Delta(H \ominus 2x)| = O(K)$. From this it follows that $|(H \oplus x)\Delta(H \ominus 2x)| = O(K)$, which implies that $3x \in H$ if $n \geq CK$ is large enough.

If γ is not an elliptic curve or an acnodal cubic then the group γ^* is isomorphic to \mathbb{R} or to $\mathbb{R} \times \mathbb{Z}/2\mathbb{Z}$, and neither of these groups has a finite subgroup of size larger than 2. □

We turn now to the consideration of sets P which are almost contained in the union of an irreducible conic σ and a line ℓ . The union $\sigma \cup \ell$ is a reducible cubic and so does not have a *bona fide* group law. There is, however, a very good substitute for one as the following proposition shows. In what follows we write $\sigma^* := \sigma \setminus (\sigma \cap \ell)$ and $\ell^* := \ell \setminus (\sigma \cap \ell)$. Note that the intersection $\sigma \cap \ell$ has size 0, 1 or 2.

Proposition 7.3 (Quasi-group law) *Suppose that σ is an irreducible conic and that ℓ is a line. Then there is an abelian group $G = G_{\sigma, \ell}$ with operation \oplus and bijective*

maps $\psi_\sigma : G \rightarrow \sigma^*, \psi_\ell : G \rightarrow \ell^*$ such that $\psi_\sigma(x), \psi_\sigma(y)$ and $\psi_\ell(z)$ are collinear precisely if $x \oplus y \oplus z = 0$. Furthermore $G_{\sigma,\ell}$ is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{R}$ if $|\sigma \cap \ell| = 2$, to \mathbb{R} if $|\sigma \cap \ell| = 1$ and to \mathbb{R}/\mathbb{Z} if $|\sigma \cap \ell| = 0$.

Proof This is certainly a known result, but it is also an easy and fun exercise to work through by hand, as we now sketch. If $|\sigma \cap \ell| = 2$, we may apply a projective transformation to move the two points of intersection to $[0, 0, 1]$ and $[0, 1, 0]$, and σ^* to the parabola $\{[a, a^2, 1] : a \in \mathbb{R}^\times\}$ and ℓ^* to $\{[0, -b, 1] : b \in \mathbb{R}^\times\}$. We note that $[a_1, a_1^2, 1], [a_2, a_2^2, 1]$ and $[0, -b, 1]$ are collinear if and only if $b = a_1 a_2$.

Consider the maps $\psi_\ell : \mathbb{Z}/2\mathbb{Z} \times \mathbb{R} \rightarrow \ell^*$ defined by $\psi_\ell(\varepsilon, x) = [0, -b, 1]$ where $b = (-1)^\varepsilon e^{-x}$ and $\psi_\sigma : \mathbb{Z}/2\mathbb{Z} \times \mathbb{R} \rightarrow \sigma^*$ defined by $\psi_\sigma(\varepsilon, x) = [a, a^2, 1]$ where $a = (-1)^\varepsilon e^x$. Then we see that the maps ψ_ℓ, ψ_σ are bijections and that the claimed collinearity property holds.

Now suppose that we are in case (ii), that is to say $|\sigma \cap \ell| = 1$. Applying a projective transformation, we may suppose that the point of intersection is $[0, 1, 0]$ and move σ^* to the parabola $\{[a, a^2, 1] : a \in \mathbb{R}\}$ and ℓ^* to the line at infinity $\{[1, -b, 0] : b \in \mathbb{R}\}$. Now note that if $[a_1, a_1^2, 1], [a_2, a_2^2, 1]$ and $[1, -b, 0]$ are distinct and collinear, then $a_1 + a_2 + b = 0$ [cf. the near-example (2.4)].

Finally suppose that σ and ℓ do not intersect. Applying a projective transformation we may map ℓ to the line at infinity $\{[\sin \pi\theta, \cos \pi\theta, 0] : \theta \in \mathbb{R}/\mathbb{Z}\}$. As σ is disjoint from ℓ it must be a compact conic section in \mathbb{R}^2 , that is to say an ellipse. By a further affine transformation we may assume that it is in fact the unit circle $\{[\cos 2\pi\theta, \sin 2\pi\theta, 1] : \theta \in \mathbb{R}/\mathbb{Z}\}$. By elementary trigonometry it may be verified that the points $[\cos 2\pi\alpha_1, \sin 2\pi\alpha_1, 1], [\cos 2\pi\alpha_2, \sin 2\pi\alpha_2, 1]$ and $[\sin \pi\beta, \cos \pi\beta, 0]$ are collinear if and only if $\alpha_1 + \alpha_2 + \beta = 0$, thus in this case the result is true with $\psi_\sigma(\theta) = [\cos 2\pi\theta, \sin 2\pi\theta, 1]$ and $\psi_\ell(\theta) = [\sin \pi\theta, \cos \pi\theta, 0]$. □

We may now derive the following consequence, analogously to Lemma 7.2.

Lemma 7.4 (Conic and line) *Suppose that $P \subset \mathbb{RP}^2$ is a set of $n \geq n_0(K)$ points, all except K of which lie on the union of an irreducible conic σ and a line ℓ . Suppose that P defines at most Kn ordinary lines, and suppose that P has $n/2 + O(K)$ points on each of σ and ℓ . Then, after a projective transformation, P differs from one of the sets $X_{n'}$ by at most $O(K)$ points.*

Proof Write $P' := P \cap (\sigma \cup \ell)$. Set $P_\sigma := P \cap \sigma^*$ and $P_\ell := P \cap \ell^*$. Then $|P_\sigma| + |P_\ell| = |P| + O(K)$, and P' spans at most $O(Kn)$ ordinary lines. Consider also the pull-backs $A := \psi_\sigma^{-1}(P_\sigma)$ and $B := \psi_\ell^{-1}(P_\ell)$, where ψ_ℓ, ψ_σ are the “quasi-group law” maps introduced in the preceding proposition. Both A and B are subsets of $G_{\sigma,\ell}$, a group for which there are three possibilities, detailed in Proposition 7.3.

The assumption about ordinary lines implies that $\ominus a_1 \ominus a_2 \in B$ for all but at most $O(Kn)$ pairs $a_1, a_2 \in A$. Applying Proposition A.5, it follows that there is a subgroup $H \leq G_\gamma$ and cosets $x \oplus H, -2x \oplus H$ such that $|A\Delta(x \oplus H)|, |B\Delta(-2x \oplus H)| = O(K)$.

If $n \geq CK$ for large enough C then it follows that $|\sigma \cap \ell| = 0$ since (with reference to the three possibilities for $G_{\sigma,\ell}$ described in Proposition 7.3) neither $\mathbb{Z}/2\mathbb{Z} \times \mathbb{R}$ nor \mathbb{R} has a finite subgroup of size larger than 2. Applying a projective transformation,

we may assume that ℓ is the line at infinity and, as in the proof of Proposition 7.3, that σ is the unit circle. We may apply a further rotation so that $x = 0$, that is to say $|A\Delta H| = O(K)$ and $|B\Delta H| = O(K)$.

All finite subgroups of \mathbb{R}/\mathbb{Z} are cyclic and so we have $H = \{j/m : j \in \{0, 1, \dots, m - 1\}\}$ for some m , that is to say H consists of the (additive) m th roots of unity. But then $\psi_\sigma(H) \cup \psi_\ell(H)$ is precisely the set $X_{n'} = X_{2m}$ described in the introduction and in the statement of Theorem 1.5. □

Putting Lemmas 7.2 and 7.4 together with the main result of the previous section, Proposition 6.14, we immediately obtain Theorem 7.1. The remainder of this section is devoted to establishing our most precise structure theorem, Theorem 1.5. Let us remind the reader that this is the same as Theorem 7.1, only the polynomial error terms $O(K^{O(1)})$ are replaced by linear errors $O(K)$. The reader interested in the proof of the Dirac–Motzkin conjecture for large n may proceed immediately to the next section, where Theorem 7.1 is already sufficient.

Proof of Theorem 1.5 The converse claim to this theorem already follows from the analysis in Sect. 2, so we focus on the forward claim. We may assume that the constant C is sufficiently large. We then apply Theorem 7.1 to obtain (after a projective transformation) that P differs by $O(K^{O(1)})$ points from one of the three examples (i), (ii), (iii) listed. Our task is to bootstrap this $O(K^{O(1)})$ error to a linear error $O(K)$.

Suppose first that case (i) holds, thus all except $O(K^{O(1)})$ points of P lie on a line ℓ . Then every point p in P that does not lie on ℓ forms at least $n - O(K^{O(1)})$ lines with a point in $P \cap \ell$. At most $O(K^{O(1)})$ of these can meet a further point in P , so each point in $P \setminus \ell$ produces at least $n - O(K^{O(1)})$ ordinary lines connecting that point with a point in $P \cap \ell$. We conclude that the number of ordinary lines is at least $(n - O(K^{O(1)}))|P \setminus \ell|$; since there are at most Kn ordinary lines, we conclude that $|P \setminus \ell| = O(K)$, and the claim follows.

Now suppose that case (ii) holds, thus P differs by $O(K^{O(1)})$ points from X_{2m} for some $m = \frac{1}{2}n - O(K^{O(1)})$. To analyse this we need the following result, essentially due to Poonen and Rubinstein [29]. □

Proposition 7.5 *Let $\Pi_n \subset \mathbb{C} \cong \mathbb{R}^2$ denote the regular n -gon consisting of the n th roots of unity. Then no point other than the origin or an element of Π_n lies on more than C lines joining pairs of vertices in Π_n , for some absolute constant C .*

Actually, in [29] it was shown that C could be taken to be 7 when one restricts attention to points *inside* the unit circle. The case of points outside the unit circle was not directly treated in that paper, but can be handled by a variant of the methods of that paper. See Appendix B for details. For the purposes of establishing Theorem 1.5, the full strength of Proposition 7.5 is unnecessary. Indeed, the more elementary weaker version established in Proposition B.2 would also suffice for this purpose.

Corollary 7.6 *Suppose that $p \in \mathbb{R}P^2$ does not lie on the line at infinity, is not an m th root of unity and is not the origin $[0,0,1]$. Then at least $2m - O(1)$ of the $2m$ lines joining p to a point of X_{2m} pass through no other point of X_{2m} .*

Proof If $x \in X_{2m}$ then we say that x is *bad* if the line $\overline{\{p, x\}}$ passes through another point $y \in X_{2m}, y \neq x$. Suppose first that x is an m^{th} root of unity. We claim that if x is *bad* then px passes through another m^{th} root of unity, different from x , or else $\overline{\{p, x\}}$ is tangent to the unit circle. If y is already an m^{th} root of unity then we are done; otherwise y is one of the m points on the line at infinity. But then the line $\overline{\{x, y\}}$ passes through another m^{th} root of unity x' unless it is tangent to the unit circle, and we have proved the claim.

This enables us to count the number of *bad* x which are m^{th} roots of unity. There are at most two coming from the possibility that $\overline{\{p, x\}}$ is tangent to the unit circle. Otherwise, $\overline{\{p, x\}}$ contains another m^{th} root of unity x' , whence p lies on the line $\overline{\{x, x'\}}$. This gives at most $O(1)$ further possibilities by Proposition 7.5.

Now suppose that x is one of the m points on the line at infinity. If $\overline{\{p, x\}}$ passes through an m^{th} root of unity y then it passes through another such root of unity y' unless $\overline{\{p, x\}}$ is tangent to the unit circle. There are at most 2 points on the line at infinity corresponding to the tangent lines, and then at most $O(1)$ corresponding to the chords $\{y, y'\}$ on which p lies, by another application of Proposition 7.5. \square

We return now to the analysis of case (ii). Let p be a point of P not on either the unit circle or the line at infinity. Then by Corollary 7.6, only $O(K^{O(1)})$ of the $n - O(K^{O(1)})$ lines connecting p with X_{2m} , also meet another element of X_{2m} . As P only differs from X_{2m} by $O(K^{O(1)})$ points, we conclude that there are $n - O(K^{O(1)})$ ordinary lines of P that connect p with an element of X_{2m} . As in case (i), this implies that there are at most $O(K)$ points of P lying outside the union of the unit circle and the line at infinity. Applying Lemma 7.4, we obtain the claim.

Finally, we consider the case (iii). By Lemma 7.2, it suffices to show that there are at most $O(K)$ points of P that do not lie on the curve E , which is either an elliptic curve or the smooth points of an acnodal singular cubic curve. By the same argument used to handle cases (i) and (ii), it then suffices to show that each point p in $P \setminus E$ generates $\gg n$ ordinary lines in P .

For this, it suffices to establish the following lemma.

Lemma 7.7 *Suppose that E is an elliptic curve or the smooth points of an acnodal singular cubic curve and that $H \oplus x$ is a coset of a finite subgroup of E of size $n > 10^4$. Then, if $p \notin E$ is a point, there are at least $n/1000$ lines through p that meet exactly one element of $H \oplus x$. \square*

Remark The constant $1/1000$ could be improved a little by our methods, but we have not bothered to perform such an optimisation here. It seems reasonable to conjecture, in analogy with the results in [29], that in fact there are only $O(1)$ lines through p that can meet three elements of a coset $x \oplus H$ of a finite subgroup of an elliptic curve, but this would seem to lie far deeper.

Proof We first exclude one degenerate case, in which E is the smooth points of an acnodal singular cubic curve, and p is the isolated (i.e. acnodal) singular point of that curve. In this case, any line through p meets exactly one point of E , and the claim is trivial. Thus, we may assume that p does not lie on the cubic curve that contains E .

Suppose the result is false. Then at least $0.999n$ of the lines joining p to $x \oplus H$ meet $x \oplus H$ in 2 or 3 points. In the former case, the line must be tangent to E . There

are at most 6 such tangents.⁷ Thus at least $0.998n$ of the lines joining p to points of $x \oplus H$ meet $x \oplus H$ in 3 points.

As a topological group, the cubic curve E is isomorphic to either \mathbb{R}/\mathbb{Z} or $\mathbb{R}/\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$. Consider all the lines through p that are tangent to E ; there are at most 6 such lines, each meeting E in at most 2 points. These (at most) 12 points partition E into no more than 13 connected open sets A_1, \dots, A_{13} (topologically, these are either arcs or closed loops), plus 12 endpoint vertices. From a continuity argument, we see that for each i one of the following statements is true:

- (i) the lines connecting p to points of A_i do not meet E again;
- (ii) there exist A_j, A_k , distinct from each other and from A_i , such that any line connecting p and a point in A_i meets E again, once at a point in A_j , and once at a point in A_k .

Suppose that i is of type (i). Then by our supposition that the lemma is false we may assume that $|A_i \cap (H \oplus x)| < 0.001n$, since all the lines from p to $A_i \cap (H \oplus x)$ contain no other point of E . By the pigeonhole principle there is some i of type (ii) with $|A_i \cap (H \oplus x)| > \frac{1}{13}(1 - 0.012)n > 0.05n > 3$. By property (ii), lines from p through A_i meet the curve E again in A_j and A_k .

Recall that for all except at most $0.002n$ elements q of $A_i \cap (H \oplus x)$, the line $\overline{\{p, q\}}$ meets A_j and A_k at elements of $H \oplus x$. It is easy to conclude from this, and similar statements for j, k , that the sizes of $A_i \cap (H \oplus x)$, $A_j \cap (H \oplus x)$ and $A_k \cap (H \oplus x)$ differ by at most $0.004n$.

Let $\phi_{ij} : A_i \rightarrow A_j$ be the map that sends a point q in A_i to the point $\overline{\{p, q\}} \cap A_j$, then ϕ_{ij} is a homeomorphism from the set A_i to the set A_j ; in particular, ϕ_{ij} is either orientation-preserving or orientation-reversing, once one places an orientation on both A_i and A_j . Furthermore, ϕ_{ij} maps all but at most $0.002n$ of the elements of $A_i \cap (H \oplus x)$ to $A_j \cap (H \oplus x)$ and vice versa. Now as H is a subgroup of E , which as an abelian topological group is either \mathbb{R}/\mathbb{Z} or $\mathbb{R}/\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$, we see that the sets $A_i \cap (H \oplus x)$, $A_j \cap (H \oplus x)$, being intersections of arcs in \mathbb{R}/\mathbb{Z} with the discrete coset $x \oplus H$, are arithmetic progressions in $x \oplus H$ with a common spacing h .

Now for all but at most $0.002n$ values of $y \in A_i \cap (H \oplus x)$, ϕ_{ij} maps y to a point of $A_j \cap (H \oplus x)$. For all but at most $1 + 0.002n$ values of $y \in A_i \cap (H \oplus x)$, ϕ_{ij} maps $y \oplus h$ to a point of $A_j \cap (H \oplus x)$. (The extra 1 comes from the fact that there is one endpoint value of y in the progression $A_i \cap (x \oplus H)$ for which $y \oplus h$ does not lie in this progression.) Of the values of y satisfying both of these statements, for all but at most $0.004n$ values we have

$$\phi_{ij}(y \oplus h) = \phi_{ij}(y) \oplus h' \tag{7.1}$$

for h' equal to either h or $\ominus h$ (depending on whether ϕ_{ij} is orientation-preserving or orientation-reversing). Thus (7.1) holds for all except at most $0.01n$ values of $y \in A_i \cap (H \oplus x)$.

⁷ The points of tangency must all lie on the intersection of E with its dual curve with respect to p , which has degree 2; see also the *Plücker Formule*.

Similarly, defining ϕ_{ik} in exactly the same way as ϕ_{ij} , we see that for all except at most $0.01n$ elements y in $A_i \cap (H \oplus x)$ we have

$$\phi_{ik}(y \oplus h) = \phi_{ik}(y) \oplus h'' \tag{7.2}$$

for h'' equal to either h or $\ominus h$.

Recalling that $|A_i \cap (x \oplus H)| > 0.05n$, we may thus find $y \in (x \oplus H) \cap A_i$ such that both (7.1) and (7.2) hold. On the other hand, as $y, \phi_{ij}(y), \phi_{ik}(y)$ are collinear, we have

$$y \oplus \phi_{ij}(y) \oplus \phi_{ik}(y) = O$$

and similarly

$$y \oplus h \oplus \phi_{ij}(y \oplus h) \oplus \phi_{ik}(y \oplus h) = O.$$

From these equations and (7.1), (7.2) we conclude that

$$h \oplus h' \oplus h'' = O.$$

Since h', h'' are equal to either h or $\ominus h$, we conclude that h has order at most 3, and so $|A_i \cap (H \oplus x)| \leq 3$. However we have already observed that $|A_i \cap (H \oplus x)| > 3$, a contradiction. □

The proof of Theorem 1.5 is now complete.

8 The Dirac–Motzkin Conjecture

The Dirac–Motzkin conjecture is the statement that, for n large, a set of P points in \mathbb{R}^2 not all lying on a line spans at least $n/2$ ordinary lines. The main result of this paper is a proof of a more precise version of this for large n , Theorem 2.2, together with a characterization of the extremal examples. We prove the even more precise Theorem 2.4, of which Theorem 2.2 is an easy consequence, in this section. We refer the reader to Sect. 2 for a precise statement of these two results and a leisurely discussion of the relevant examples.

Suppose that P spans at most n ordinary lines and that P is not collinear. We may apply our main structure theorem, Theorem 1.5, to conclude that P differs in $O(1)$ points from one of three examples: points on a line, a set $X_{n'}$, and a coset of a subgroup of an irreducible cubic curve (Sylvester-type example). In fact, the weaker and rather easier Theorem 7.1 suffices for this purpose.

It is obvious that the first type of set spans at least $n - O(1)$ ordinary lines. Sets close to a Sylvester example are also relatively easy to handle.

Lemma 8.1 *Suppose that $P \subset \mathbb{RP}^2$ differs in K points from a coset $H \oplus x$ of a subgroup H of some irreducible cubic curve, where $3x = x \oplus x \oplus x \in H$. Then P spans at least $n - O(K)$ ordinary lines.*

Proof Write $h_0 := 3x$, thus $h_0 \in H$. For every $h \in H$, the line joining $h \oplus x$ and $(-2h \oplus h_0) \oplus x$ is tangent to γ at $h \oplus x$, since $(h \oplus x) \oplus (h \oplus x) \oplus (-2h \oplus h_0 \oplus x) = 0$. Therefore it is an ordinary line unless $3h \oplus h_0 = 0$, in which case the points $h \oplus x$ and $-2h \oplus h_0 \oplus x$ coincide. Thus the only points of $H \oplus x$ not belonging to an ordinary line spanned by $H \oplus x$ correspond to the points of H with $3h \oplus h_0 = 0$. Since H is isomorphic to a subgroup of either \mathbb{R}/\mathbb{Z} or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$, there are no more than 3 of these. It follows immediately that any set formed by removing at most K points of $H \oplus x$ has at least $n - O(K)$ ordinary lines, and these are all tangent lines to γ . No point in the plane lies on more than 6 tangent lines to γ , and so the addition of a point destroys no more than 6 of our $n - O(K)$ ordinary lines. It follows that P itself spans at least $n - O(K)$ ordinary lines, as we wanted to prove. \square

Combining this lemma with the remarks just preceding it, we have now established the existence of an absolute constant C such that a set of n points, not all on a line, and spanning at most $n - C$ ordinary lines, differs in $O(1)$ points from a set X_{2m} consisting of the m th roots of unity plus m corresponding points on the line at infinity. Now the m tangents to the unit circle at roots of unity pass through only one other point of X_{2m} , and so X_{2m} has m ordinary lines. Furthermore, since each point not on the unit circle can be incident to at most two such tangent lines, the addition/deletion of $O(1)$ points does not affect more than $O(1)$ of these lines. This already establishes a weak version of the Dirac–Motzkin conjecture: every non-collinear set of n points spans at least $n/2 - O(1)$ ordinary lines.

To prove Theorem 2.4, a much more precise result, we must analyse configurations close to X_{2m} more carefully. What is needed is precisely the following result which, together with what we have already said in this section, completes the proof of Theorem 2.4. Recall from Sect. 2 the examples of Böröczky.

Proposition 8.2 *There is an absolute constant C such that the following is true. Suppose that P differs from X_{2m} in at most K points, and that P spans at most $2m - CK$ ordinary lines. Then P is a Böröczky example or a near-Böröczky example.*

We now prove this proposition. Suppose that P differs from X_{2m} in at most K points. Suppose first of all that P contains a point p outside X_{2m} , that p does not lie on the line at infinity and that $p \neq [0, 0, 1]$. Then by Corollary 7.6, the m tangent lines at the unit circle, as well as at least $m - O(1)$ of the lines $\overline{\{p, x\}}$ connecting p to a point $x \in X_{2m}$, pass through precisely 2 points of $X_{2m} \cup \{p\}$. It is clear that the addition/deletion of K points *other than* p cannot add or delete points on more than $O(K)$ of these lines, and so P spans $2m - O(K)$ ordinary lines.

Now suppose that P contains an additional point p on the line at infinity. Then the m tangent lines to the m th roots of unity, as well as the at least $2m - 2$ lines $\overline{\{p, x\}}$, x an m th root of unity, which are not tangent to the unit circle contain precisely two points of $X_{2m} \cup \{p\}$. Once again the addition/deletion of K points *other than* p cannot add or delete points on more than $O(K)$ of these lines, and so again P spans $3m - O(K)$ ordinary lines.

We have now reduced matters to the case $P \subset X_{2m} \cup [0, 0, 1]$. Starting from X_{2m} , the omission of a point or the addition of $[0, 0, 1]$ creates a certain number of new lines with precisely two points, and of course no point other than $[0, 0, 1]$ or the omitted

point is on more than one of these new lines. By inspection in any case there are always at least $m/2 - O(1)$ of these lines, and so there are at least $2m - O(K)$ ordinary lines unless we do *at most one* of the operations of adding $[0, 0, 1]$ or removing a point of X_{2m} . At this point a short inspection of the possibilities leads to the conclusion that the Böröczky examples and the near-Böröczky examples are the only ones which do not have at least $3m - O(1)$ ordinary lines. This, at last, concludes the proof of Theorem 2.4. \square

Remark We relied on Corollary 7.6, which depended on the result of Poonen and Rubinstein [29]. For the purposes of proving the Dirac–Motzkin conjecture for large n , the somewhat easier Proposition B.2 is sufficient.

9 The Orchard Problem

In this section we establish Theorem 1.3, the statement that a set of n points in the plane contains no more than $\lfloor \frac{1}{6}n(n - 3) \rfloor + 1$ 3-rich lines when n is sufficiently large. The sharpness of this bound was established in Proposition 2.6.

If N_k is the number of lines containing precisely k points of P then, by double-counting pairs of points in P , we have

$$\sum_{k \geq 2} \binom{k}{2} N_k = \binom{n}{2}. \tag{9.1}$$

From this it follows that if $N_3 > \lfloor \frac{1}{6}n(n - 3) \rfloor + 1$ then

$$N_2 + \sum_{k \geq 4} \binom{k}{2} N_k \leq n, \tag{9.2}$$

from which we conclude that N_2 , the number of ordinary lines spanned by P , is at most n . Furthermore no line contains more than $O(\sqrt{n})$ points.

We may now apply Theorem 1.5, our structure theorem for sets with few ordinary lines. Since no line contains more than $O(\sqrt{n})$ points of P we see that in fact only option (iii) of that theorem can occur, that is to say P differs in $O(1)$ points from a coset $H \oplus x$, $3x \in H$, of a subgroup H of some irreducible cubic curve γ , which is either an elliptic curve or (the smooth points of) an acnodal singular curve. The rest of the analysis is straightforward but a little tedious.

Suppose that $3x = h_0$. As in the proof of Lemma 8.1, the tangent line to γ at $h \oplus x$ meets $H \oplus x$ in the point $(-2h \oplus h_0) \oplus x$, which is distinct from the first point unless $3h \oplus h_0 = 0$. There are at most $O(1)$ of these. In creating P from $H \oplus x$ by the addition/deletion of $O(1)$ points, at most $O(1)$ of these lines are affected.

Since P spans at most n ordinary lines, it follows that P contains only $O(1)$ ordinary lines other than these tangent lines. Furthermore, since we now know that P contains at least $n + O(1)$ ordinary lines, that is to say $N_2 = n + O(1)$, we conclude from (9.2) that $N_4 = O(1)$. We are going to conclude that $P = H \oplus x$, a statement whose proof we divide into three parts.

Claim 1 There is no point of P off the curve γ . If p is such a point, all except $O(1)$ of the lines joining p to points of $P \cap \gamma$ must contain precisely two points of $P \cap \gamma$, or else there would be too many lines containing p with 2 or 4 points of P . Note that this cannot happen if γ is the smooth points of an acnodal singular cubic curve and p is the isolated singular point, since every line through p meets at most one point of γ ; thus p lies outside of the cubic curve containing γ . Consider the lines ℓ from p to $P \cap \gamma$ which are not tangent to γ and which contain precisely two points of $P \cap \gamma$ and precisely two points of $H \oplus x$, these points being the same. Since $P \cap \gamma$ differs from $H \oplus x$ in $O(1)$ points, all except $O(1)$ of the lines from p have this property. But any line not tangent to γ and containing the two points $h \oplus x$ and $h' \oplus x$ also contains $-(h \oplus h' \oplus h_0) \oplus x$, a third point of $H \oplus x$. This is a contradiction.

Claim 2 There is no point of P outside the set $H \oplus x$. Suppose that $k \oplus x$ is such a point. Then if $h \in H$, the line joining $k \oplus x$ and $h \oplus x$ meets γ again at $-(k \oplus h \oplus h_0) \oplus x$, which is not a point of $H \oplus x$. This point can thus only lie in P for $O(1)$ values of h , and hence there are $n - O(1)$ ordinary lines of P emanating from $k \oplus x$. In addition to the $n - O(1)$ tangent lines, this gives at least $2n - O(1)$ ordinary lines in P , a contradiction.

Claim 3 P contains all of $H \oplus x$. Suppose that $h_* \oplus x$ is a point of $H \oplus x$ not contained in P . For all except $O(1)$ values of h , the points $h \oplus x$ and $-(h \oplus h_* \oplus h_0) \oplus x$ lie in P , and the line joining $h_* \oplus x$ to them is not tangent to γ . All such lines then contain precisely two points of P , and once again we obtain $n - O(1)$ ordinary lines to add to the $n - O(1)$ tangent lines we already have. Once again a contradiction ensues.

We have now shown that if P is a set of n points in the plane with N_3 , the number of lines in P spanning precisely 3 points, satisfying $N_3 > \lfloor \frac{1}{6}n(n-3) \rfloor + 1$, then P is a coset $H \oplus x$ on γ , an elliptic curve or the smooth points of an acnodal cubic, with $3x \in H$. But by Proposition 2.6 we have $N_3 \leq \lfloor \frac{1}{6}n(n-3) \rfloor + 1$ in any such case, and we are done.

Remarks Note that, we have in fact classified (for large n) the optimal configurations in the orchard problem as coming from cosets in elliptic curves or acnodal cubics. We note that nothing like the full force of Theorem 1.5 is required for the orchard problem (as opposed to the Dirac–Motzkin conjecture). Once the much weaker Proposition 5.3 is established, we can immediately rule out possibilities (ii) and (iii) of that proposition and hence do away with all of the material in Sect. 6 and some of the material in Sect. 7 too.

Acknowledgments The authors are greatly indebted to Luke Alexander Betts for indicating to us the proof of Proposition 6.4. We are also grateful to Noga Alon, Boris Bukh, Frando Mariacci, Bjorn Poonen and Jozsef Solymosi for helpful comments and conversations, and to Frank de Zeeuw for pointing out the case of the acnodal singular cubic curve, which was not treated properly in the first draft of this paper. The first author thanks UCLA for providing entertaining working conditions for some of the period during which this work was completed, and the second author similarly thanks the Newton Institute for hospitality when another portion of this work was completed. Finally, we would like to thank the anonymous referee for an extremely thorough reading of the paper. The first author was supported by an ERC Starting Investigators Grant. The second author was partially supported by a Simons Investigator award from the Simons Foundation and by NSF grant DMS-0649473.

Appendix A: Some Tools from Additive Combinatorics

In this section, we collect some more-or-less standard tools from additive combinatorics used in Sects. 6 and 7.

If A, B are two sets in some abelian group, and if $\Gamma \subset A \times B$ is a set of pairs, we write $A +_{\Gamma} B := \{a + b : (a, b) \in \Gamma\}$. The next result is known as the Balog–Szemerédi–Gowers theorem. The precise form we use is a variant of Gowers’s version [19, Proposition 12] due to Sudakov, Szemerédi and Vu [31, Theorem 4.1].

Theorem A.1 (Balog–Szemerédi–Gowers) *Suppose that A, B are two sets in an abelian group, both of size at most n . Suppose that $\Gamma \subset A \times B$ is a set (which may be thought of as a bipartite graph) with $|\Gamma| \geq n^2/K$. Suppose that $|A +_{\Gamma} B| \leq K'n$. Then there are sets $A' \subset A, B' \subset B$ with $|A'| \geq \frac{n}{4K}, |B'| \geq \frac{n}{16K^2}$ such that $|A' + B'| \leq 2^{12}(K')^3 K^5 n$.*

Proof See [31, Theorem 4.1]. In the statement of that result A and B are both supposed to have size n , but it is easy to see that the proof works under the assumption that they both have size at most n , for instance by adding dummy elements to A or B (enlarging the group G if necessary) while keeping Γ unchanged. \square

On several occasions in Sect. 6 we will apply the preceding theorem together with Ruzsa’s triangle inequality (see e.g. [35, Lemma 2.6]), which states that $|U||V - W| \leq |U - V||U - W|$ for any sets U, V, W in an abelian group (in fact, the group does not even need to be abelian). Let us record, as a corollary, the result of doing this in the particular context we need.

Corollary A.2 *Suppose that A, B are two sets in an abelian group, both of size at most n . Suppose that $\Gamma \subset A \times B$ is a set with $|\Gamma| \geq \delta n^2$ for which $|A +_{\Gamma} B| \leq n$. Then there are sets $A' \subset A, B' \subset B$ with $|A'|, |B'| \gg \delta^2 n$ such that $|A' - A'|, |B' - B'| \ll \delta^{-11} n$.*

Proof This follows immediately from the preceding lemma and the Ruzsa triangle inequality.

The following is a “robust” version of the elementary sumset estimate $|U + V| \geq |U| + |V| - 1$.

Lemma A.3 *Let $U, V \subset \mathbb{R}$ be sets of size r and s respectively. Suppose that $\Gamma \subset U \times V$ has cardinality at least $(1 - \delta)rs$. Then $|U +_{\Gamma} V| \geq r + s - 2 - 2\sqrt{2\delta}rs$.*

Proof Suppose that $U = \{u_1, \dots, u_r\}$ with $u_1 < \dots < u_r$, and $V = \{v_1, \dots, v_s\}$ with $v_1 < \dots < v_s$. For any $1 \leq k \leq \min(r, s)$ we have

$$u_1 + v_k < u_2 + v_k < \dots < u_{r-k} + v_k < u_{r-k} + v_{k+1} < \dots < u_{r-k} + v_s,$$

giving $r + s - 2k$ distinct elements of $U + V$. As k varies, no pair (u_i, v_j) appears in this listing more than twice. Thus by the pigeonhole principle there is, for any choice

of positive integer k_0 , some $k \leq k_0$ such that at most $2\delta rs/k_0$ elements of this listing come from pairs (u_i, v_j) not lying in Γ . It follows that

$$|U +_{\Gamma} V| \geq r + s - 2k_0 - \frac{2\delta rs}{k_0}.$$

Choosing $k_0 := \lceil \sqrt{\delta rs} \rceil$ confirms the result. □

We actually need a variant of this result for subsets of the *multiplicative* group \mathbb{R}^* . If $U, V \subset \mathbb{R}^*$ and if $\Gamma \subseteq U \times V$ then we write $U \cdot_{\Gamma} V = \{uv : (u, v) \in \Gamma\}$.

Lemma A.4 *Let $U, V \subset \mathbb{R}^*$ be sets of size r and s respectively. Suppose that $\Gamma \subset U \times V$ has cardinality at least $(1 - \delta)rs$. Then $|U \cdot_{\Gamma} V| \geq r + s - 4 - 2\sqrt{2\delta rs}$.*

Proof As an additive group, \mathbb{R}^* is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{R}$. By abuse of notation, we identify U and V with subsets of this additive group and use additive notation. Define $U_0 = (\{0\} \times \mathbb{R}) \cap U, U_1 = (\{1\} \times \mathbb{R}) \cap U, V_0 = (\{0\} \times \mathbb{R}) \cap V$ and $V_1 = (\{1\} \times \mathbb{R}) \cap V$. Write $r_0 = |U_0|, r_1 = |U_1|, s_0 = |V_0|$ and $s_1 = |V_1|$. Suppose that $\Gamma \cap (U_i \times V_j)$ has $\delta_{i,j}r_i s_j$ edges; then

$$\delta_{0,0}r_0s_0 + \delta_{0,1}r_0s_1 + \delta_{1,0}r_1s_0 + \delta_{1,1}r_1s_1 = \delta rs. \tag{10.1}$$

Clearly

$$(U +_{\Gamma} V) \cap (\{0\} \times \mathbb{R}) \supset U_0 +_{\Gamma} V_0, \quad U_1 +_{\Gamma} V_1$$

and

$$(U +_{\Gamma} V) \cap (\{1\} \times \mathbb{R}) \supset U_1 +_{\Gamma} V_0, \quad U_0 +_{\Gamma} V_1.$$

Therefore by the preceding lemma

$$\begin{aligned} |U +_{\Gamma} V| &\geq \max(r_0 + s_0 - 2\sqrt{2\delta_{0,0}r_0s_0}, r_1 + s_1 - 2\sqrt{2\delta_{1,1}r_1s_1}) \\ &\quad + \max(r_0 + s_1 - 2\sqrt{2\delta_{0,1}r_0s_1}, r_1 + s_0 - 2\sqrt{2\delta_{1,0}r_1s_0}) - 4 \\ &\geq r_0 + s_0 + r_1 + s_1 - \sum_{i,j} \sqrt{2\delta_{i,j}r_i s_j} - 4. \end{aligned}$$

Using the inequality

$$\sqrt{x} + \sqrt{y} + \sqrt{z} + \sqrt{w} \leq 2\sqrt{x + y + z + w}$$

(easily established using Cauchy–Schwarz) together with (10.1), we obtain

$$|U +_{\Gamma} V| \geq r + s - 2\sqrt{2\delta rs} - 4,$$

as claimed. □

The following result was used heavily in Sect. 7. It is of a fairly standard type and will be of no surprise to experts in additive combinatorics, but we do not know of a convenient reference.

Proposition A.5 *Suppose that A, B, C are three subsets of some abelian group G , all of cardinality within K of n , where $K \leq \varepsilon n$ for some absolute constant $\varepsilon > 0$. Suppose that there are at most Kn pairs $(a, b) \in A \times B$ for which $a + b \notin C$. Then there is a subgroup $H \leq G$ and cosets $x + H, y + H$ such that $|A\Delta(x + H)|, |B\Delta(y + H)|, |C\Delta(x + y + H)| \leq 7K$.*

As remarked in Sect. 7, results of this general type are quite familiar to additive combinatorialists and are of the general form “an almost-group is close to a group”. We supply a complete proof here for the convenience of the reader. Variants of it are possible. For the most part the ideas are due to Kneser [23,24], Freiman [17] and Fournier [16].

We first note that it is enough to prove the following weaker proposition, which may then be “cleaned up” to give the stated result.

Proposition A.6 *Let ε be a positive quantity, less than some absolute constant. Suppose that A, B, C are three subsets of some abelian group G , all of size within εn of n . Suppose that there at most εn^2 pairs $(a, b) \in A \times B$ for which $a + b \notin C$. Then there is a subgroup $H \leq G$ and cosets $x + H, y + H$ such that $|A\Delta(x + H)|, |B\Delta(y + H)|, |C\Delta(x + y + H)| \leq \varepsilon'n$, where ε' can be taken to be $O(\varepsilon^c)$ for some absolute constant $c > 0$.*

Let us deduce Proposition A.5 from this. Let A, B and C be as in the hypotheses of that proposition. Provided that ε is small enough, Proposition A.6 applies and we conclude that there is a subgroup $H \leq G$ and cosets $x + H, y + H$ such that $|A\Delta(x + H)|, |B\Delta(y + H)|, |C\Delta(x + y + H)| \leq \varepsilon'n$ with $\varepsilon' = O(\varepsilon^c)$. By translating A and B if necessary we may assume without loss of generality that $x = y = 0$.

Suppose that $A = (H \setminus X) \cup X', B = (H \setminus Y) \cup Y'$ and $C = (H \setminus Z) \cup Z'$, with $X, Y, Z \subset H, X', Y', Z'$ disjoint from H and all of X, X', Y, Y', Z, Z' having cardinality at most $\varepsilon'n$. Now if $a \in X'$ then the elements $a + b, b \in H \setminus X$, are all distinct and none of them lie in H . If such an element $a + b$ lies in C , it must therefore lie in Z' . Thus if $a \in X'$ then there are at least $|H| - |X| - |Z'| \geq (1 - 4\varepsilon')n > \frac{1}{2}n$ elements $b \in H \setminus X$ for which $a + b \notin C$. By assumption it follows that $\frac{1}{2}n|X'| \leq Kn$, which implies that $|X'| \leq 2K$. Similarly $|Y'| \leq 2K$.

Now note that, since $|X|, |Y| < \frac{1}{8}|H|$, every element of H has at least $\frac{3}{4}|H| \geq \frac{1}{2}n$ representations as a sum $a + b$. Indeed if $h \in H$ then by we have $|(h - (H \setminus X)) \cap (H \setminus Y)| > \frac{3}{4}|H|$ by the pigeonhole principle. It follows that if we pass to a subset of these sums by removing all sums $a + b$ with (a, b) lying in a set of size at most Kn , at least $|H| - 2K$ elements of H are still represented. By assumption, the set C contains a set of this form, and it follows that $|Z| \leq 2K$.

We have now demonstrated the inequalities

$$|A|, |B| \leq |H| + 2K, \quad |C| \geq |H| - 2K.$$

Since the sizes of A, B and C differ by at most K , we must in fact have

$$|H| - 3K \leq |A|, |B|, |C| \leq |H| + 3K.$$

This allows us to conclude that $|X|, |Y|, |Z'| \leq 5K$. Proposition A.5 follows immediately. □

We turn now to the task of proving Proposition A.6. We require the following result, which could be deduced from results of Kneser [23, 24] and Freiman [17].

Lemma A.7 *Let $\varepsilon < \frac{1}{60}$. Suppose that A is a subset of an abelian group G with $|A| = n$, and suppose that $|A - A| \leq (1 + \varepsilon)n$. Then there is a subgroup $H \leq G$ and a coset $x + H$ such that $|A \Delta (x + H)| \leq 6\varepsilon n$.*

Proof This is basically the argument of Fournier [16]. Write (cf. [35]) $\text{Sym}_\alpha(A)$ for the set of all d which have at least αn representations as $a_1 - a_2, a_1, a_2 \in A$. Note that $\text{Sym}_{1-\delta_1}(A) + \text{Sym}_{1-\delta_2}(A) \subset \text{Sym}_{1-\delta_1-\delta_2}(A)$, and note also that $|\text{Sym}_{5/6}(A)| \geq (1 - 5\varepsilon)n \geq \frac{11}{12}n$. This follows from double-counting pairs $(a_1, a_2) \in A^2$: we have

$$\begin{aligned} n^2 = |A|^2 &= \sum_{d \in A-A} |\{(a_1, a_2) : a_1, a_2 \in A, a_1 - a_2 = d\}| \\ &\leq |\text{Sym}_{5/6}(A)||A| + \frac{5}{6}|(A - A) \setminus \text{Sym}_{5/6}(A)||A| \\ &\leq \frac{1}{6}|\text{Sym}_{5/6}(A)|n + \frac{5}{6}(1 + \varepsilon)n^2. \end{aligned}$$

We claim that $H = \text{Sym}_{2/3}(A)$ is a group. Certainly $0 \in H$, and $H + H \subset \text{Sym}_{1/3}(A)$, so all we need do is check that $\text{Sym}_{2/3}(A) = \text{Sym}_{1/3}(A)$. Suppose that $d \in \text{Sym}_{1/3}(A)$. Then $d = a_1 - a_2$ in at least $\frac{1}{3}n$ ways. If $t \in \text{Sym}_{5/6}(A)$ then $t = a'_1 - a'_2$ in at least $\frac{5}{6}n$ ways. For at least $\frac{1}{6}n$ of these we will have $a'_1 = a_2$ for some a_2 such that $d = a_1 - a_2$, and thus $d + t = (a_1 - a_2) + (a'_1 - a'_2) = a_1 - a'_2 \in A - A$. That is, $(d + \text{Sym}_{5/6}(A)) \cap (A - A) \geq \frac{1}{6}n$. In particular, $d + \text{Sym}_{5/6}(A)$ intersects $\text{Sym}_{5/6}(A)$ (which has size at least $\frac{11}{12}n$) and therefore $d \in \text{Sym}_{5/6}(A) - \text{Sym}_{5/6}(A) \subset \text{Sym}_{2/3}(A)$, as required.

To see that A is close to a coset of H , note that $a_1 - a_2 \in H$ for all but at most $|(A - A) \setminus \text{Sym}_{2/3}(A)||A| \leq 6\varepsilon n^2$ of the pairs $(a_1, a_2) \in A$. In particular there is some $x = a_2$ such that $a_1 \in x + H$ for all but at most $6\varepsilon n$ values of $a_1 \in A$. □

We also need the following “99%” version of the Balog–Szemerédi–Gowers theorem.

Lemma A.8 *Suppose that A and B are sets in some abelian group, and that $\Gamma \subset A \times B$ is a set with $|\Gamma| \geq (1 - \varepsilon)|A||B|$. Suppose that $|A +_\Gamma B| \leq (1 + \varepsilon)|A|^{1/2}|B|^{1/2}$. Then there are sets $A' \subset A$ and $B' \subset B$ with $|A'|/|A|, |B'|/|B| \geq 1 - \varepsilon'$ such that $|A' - B'| \leq (1 + \varepsilon')|A|^{1/2}|B|^{1/2}$. We can take $\varepsilon' = O(\varepsilon^c)$ for some $c > 0$.*

Proof This follows from [35, Theorem 2.31]. If one wanted instead the conclusion $|A' + B'| \leq (1 + \varepsilon')|A|^{1/2}|B|^{1/2}$ (for comparison with Theorem A.1) then one could additionally apply [35, Proposition 2.27]. \square

With these lemmas in hand, we can conclude the proof of Proposition A.6. In what follows $\varepsilon_1, \varepsilon_2, \dots$ are all quantities bounded by $O(\varepsilon^{O(1)})$. Explicit dependencies could be given if desired, but this would require Lemma A.8 to be made explicit. With the hypotheses as in Proposition A.6, first apply Lemma A.8 to conclude that there are sets A', B' with $|A\Delta A'|, |B\Delta B'| \leq \varepsilon_1 n$ and $|A' - B'| \leq (1 + \varepsilon_2)n$. Applying the Ruzsa triangle inequality we obtain $|A' - A'|, |B' - B'| \leq (1 + \varepsilon_3)n$. By Lemma A.7 (assuming ε is sufficiently small) there are subgroups H, H' and cosets $x + H, y + H'$ such that $|A\Delta(x + H)|, |B\Delta(y + H')| \leq \varepsilon_4 n$. In particular $(1 - \varepsilon_5)n \leq |H|, |H'| \leq (1 + \varepsilon_5)n$.

Now, by assumption, for all except εn^2 pairs $(a, b) \in A \times B, a + b$ lies in a set C of size at most $(1 + \varepsilon)n$. It follows easily that for all except $\varepsilon_6 n^2$ pairs $(h, h') \in H \times H', h + h'$ lies in a set of size $(1 + \varepsilon_7)n$. We claim that this forces $H = H'$. To see this, note that the assumption easily implies that there are at least $(1 - \varepsilon_8)n^3$ additive quadruples $h_1 - h_2 = h'_1 - h'_2$, and so for all but $\varepsilon_9 n^2$ pairs $(h_1, h_2) \in H$ we have $h_1 - h_2 = h'_1 - h'_2 \in H'$. This implies that all but $\varepsilon_{10} n$ elements of H lie in H' , and hence $|H\Delta H'| \leq \varepsilon_{11} n$ and so $|H \cap H'| \geq (1 - \varepsilon_{12})n > \frac{1}{2}n$, provided ε is sufficiently small. Invoking Lagrange’s theorem (the order of a subgroup divides the order of the group), it follows that in fact $H = H \cap H' = H'$, as claimed.

Finally, note that since A occupies at least $7|H|/8$ of $x + H$, and B at least $7|H|/8$ of $y + H$, every element of $x + y + H$ is a sum $a + b$ in at least $3|H|/4 > n/2$ ways. It follows that C must contain all but at most $\varepsilon_{13} n$ of the elements of $x + y + H$, and this concludes the proof.

The next result is due to Elekes et al. [14].

Proposition A.9 *Let $A \subset \mathbb{R}$ be a set of cardinality n , and suppose that there are $x_1 < \dots < x_{10}$ such that $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined except possibly at x_1, \dots, x_{10} and is strictly concave or convex on each open interval (x_i, x_{i+1}) . Then either $|A - A|$ or $|f(A) - f(A)|$ has cardinality at least $cn^{5/4}$ for some absolute constant $c > 0$.*

Proof Suppose first of all that f is strictly convex or concave on an interval containing A . Then by [14, Corollary 3.1] we have $|A - A||f(A) - f(A)| \gg n^{5/2}$, and so either $|A - A|$ or $|f(A) - f(A)|$ has cardinality at least $cn^{5/4}$. The proposition follows by applying this to the largest of the sets $A_i := A \cap (x_i, x_{i+1})$. \square

Appendix B: Intersections of Lines Through Roots of Unity

In this section, we establish Proposition 7.5. Our arguments will be a crude variant of those used in [29].

It will be convenient to identify the plane \mathbb{R}^2 with the complex numbers \mathbb{C} . Let Π_n be the n th roots of unity, and suppose that p is a point other than the origin or an element of Π_n which is incident to m lines ℓ_1, \dots, ℓ_m , each of which pass through two points $e^{2\pi i\beta_j}, e^{2\pi i\gamma_j}$ of Π_n , where $0 \leq \beta_1, \dots, \beta_m, \gamma_1, \dots, \gamma_m < 1$ are distinct rationals with denominator n . Our objective is to show that $m = O(1)$.

We claim the identity

$$\frac{|p - e^{2\pi i\beta_j}|}{|p - e^{2\pi i\beta_k}|} = \frac{|e^{2\pi i\beta_j} - e^{2\pi i\gamma_k}|}{|e^{2\pi i\beta_k} - e^{2\pi i\gamma_j}|}$$

for any distinct $1 \leq j, k \leq m$. Indeed, from elementary trigonometry we see that $(p, e^{2\pi i\beta_j}, e^{2\pi i\gamma_k})$ and $(p, e^{2\pi i\beta_k}, e^{2\pi i\gamma_j})$ form a pair of similar triangles, regardless of the relative ordering between the points involved. The right-hand side can be simplified as

$$\frac{|\sin(\pi(\beta_j - \gamma_k))|}{|\sin(\pi(\beta_k - \gamma_j))|}.$$

We conclude that

$$\frac{|\sin(\pi(\beta_j - \gamma_k))| |\sin(\pi(\beta_k - \gamma_l))| |\sin(\pi(\beta_l - \gamma_j))|}{|\sin(\pi(\beta_k - \gamma_j))| |\sin(\pi(\beta_l - \gamma_k))| |\sin(\pi(\beta_j - \gamma_l))|} = 1$$

for any distinct $1 \leq i, j, k \leq m$, and thus

$$\begin{aligned} &\sin(\pi(\beta_j - \gamma_k)) \sin(\pi(\beta_k - \gamma_l)) \sin(\pi(\beta_l - \gamma_j)) \\ &= \pm \sin(\pi(\beta_k - \gamma_j)) \sin(\pi(\beta_l - \gamma_k)) \sin(\pi(\beta_j - \gamma_l)) \end{aligned}$$

for some choice of sign \pm . Actually, we claim that the sign here is always given by the $+$ sign. To see this, let us temporarily forget that the $e^{2\pi i\beta_j}, e^{2\pi i\gamma_k}$ were constrained to be roots of unity, and that p was assumed not to take values at the origin or at infinity (since we have not yet actually used these hypotheses). We first observe that the sign does not change if we shift any of the β_j or γ_k by an integer, so we may assume that these phases take values in \mathbb{R}/\mathbb{Z} rather than $[0, 1)$. Then we observe that the sign is stable with respect to continuous perturbations of the β_j, γ_k and p , so long as no two phases cross each other, and that $e^{2\pi i\beta_j}, e^{2\pi i\gamma_j}, p$ remain collinear for all j . From this we may reduce to the case when p is at the origin (so that $\beta_j = \gamma_j + 1/2$ for all j) or at, say, $[1, 0, 0]$ (so that $\beta_j = 1/2 - \gamma_j$ for all j), and the sign is easily verified in these cases.

Expanding out $\sin x$ as $(e^{ix} - e^{-ix})/2i$, we conclude that

$$\begin{aligned} &(e^{\pi i(\beta_j - \gamma_k)} - e^{\pi i(\gamma_k - \beta_j)})(e^{\pi i(\beta_k - \gamma_l)} - e^{\pi i(\gamma_l - \beta_k)})(e^{\pi i(\beta_l - \gamma_j)} - e^{\pi i(\gamma_j - \beta_l)}) \\ &= (e^{\pi i(\beta_k - \gamma_j)} - e^{\pi i(\gamma_j - \beta_k)})(e^{\pi i(\beta_l - \gamma_k)} - e^{\pi i(\gamma_k - \beta_l)})(e^{\pi i(\beta_j - \gamma_k)} - e^{\pi i(\gamma_k - \beta_j)}). \end{aligned}$$

Multiplying out both sides, and cancelling the common terms

$$\pm e^{\pm\pi i(\beta_j + \beta_k + \beta_l - \gamma_j - \gamma_k - \gamma_l)}$$

appearing on both sides, one arrives at an identity of the form

$$\sum_{r=1}^{12} \epsilon_r e^{\pi i \alpha_{r;j,k,l}} = 0, \tag{11.1}$$

where the $\epsilon_r = \pm 1$ are signs depending only on r , and $\alpha_{r;j,k,l}$ are twelve linear combinations of $\beta_j, \beta_k, \beta_l, \gamma_j, \gamma_k, \gamma_l$, each of the form

$$\alpha_{r;j,k,l} = \pm \beta_j \pm \beta_k \pm \beta_l \pm \gamma_j \pm \gamma_k \pm \gamma_l$$

where the six signs \pm do not need to be equal, but depend only on the index r . These signs can of course be worked out explicitly, but we will not need to do so here, save to note that the linear forms $\alpha_{r;j,k,l}$ are all distinct in r , thus $\alpha_{r;j,k,l} - \alpha_{r';j,k,l}$ is a non-trivial combination of $\beta_j, \beta_k, \beta_l, \gamma_j, \gamma_k, \gamma_l$ whenever r, r' are distinct.

Now we reinstate the hypothesis that the $e^{2\pi i \beta_j}, e^{2\pi i \gamma_k}$ are n th roots of unity, which ensures that the $\epsilon_r e^{\pi i \alpha_{r;j,k,l}}$ are also n th roots of unity. The sets of twelve n th roots of unity that sum to zero were completely classified in [29, Theorem 3]. The exact classification is somewhat messy, but we only require the following qualitative consequence of it.

Proposition B.1 *There exists a finite set S of roots of unity with the property that whenever $e^{2\pi i \alpha_1}, \dots, e^{2\pi i \alpha_{12}}$ are roots of unity with $\sum_{r=1}^{12} e^{2\pi i \alpha_r} = 0$, then one has $e^{2\pi i (\alpha_r - \alpha_{r'})} \in S$ for some $1 \leq r < r' \leq 12$.*

Indeed, one can take S to be the ratios of the roots of unity arising in the minimal relations of weight up to 12 that were classified in [29, Theorem 3], the key point being that there were only finitely many (107, to be precise) such relations up to rotation. In fact, one can take S to consist of 30030th roots of unity if desired.

Applying this proposition, we conclude that for any distinct i, j, k, l with $1 \leq i, j, k, l \leq m$, one has

$$\pm e^{\pi i (\alpha_{r;j,k,l} - \alpha_{r';j,k,l})} \in S$$

for some r, r' with $1 \leq r < r' \leq 12$ and some choice of sign \pm . Applying the pigeonhole principle, we conclude (for m large enough) that there exist r, r' with $1 \leq r < r' \leq 12$ and a phase θ such that

$$\alpha_{r;j,k,l} - \alpha_{r';j,k,l} = \theta$$

for $\gg m^3$ triples of distinct $1 \leq j, k, l \leq m$.

Fix r, r', θ as above. As mentioned earlier, $\alpha_{r;j,k,l} - \alpha_{r';j,k,l}$ is a non-trivial linear form in the $\beta_j, \beta_k, \beta_l, \gamma_j, \gamma_k, \gamma_l$. By symmetry, we may then assume that at least one of the β_j, γ_j coefficients in this form are non-zero. As these coefficients lie in $\{-2, 0, +2\}$, we may thus write

$$\alpha_{r;j,k,l} - \alpha_{r';j,k,l} = 2a\beta_j + 2b\gamma_j + c_{k,l}$$

for some coefficients $a, b \in \{-1, 0, +1\}$ not both zero, and some phases $c_{k,l}$ independent of j . Note that these coefficients are equal to $-2, 0$, or $+2$. Pigeonholing in the k, l , we may then find distinct $1 \leq k, l \leq m$ and a phase θ' such that

$$a\beta_j + b\gamma_j = \theta'$$

for $\gg m$ values of j . But from elementary trigonometry, and the hypothesis that p is not at the origin (or at infinity) we see that as $\beta, \gamma \in [0, 1)$ range over the distinct phases for which $e^{2\pi i\beta}, e^{2\pi i\gamma}$ are concurrent with p , the phase $a\beta + b\gamma$ can take on any specific value θ' at most $O(1)$ times, and so $m = O(1)$ as desired. This concludes the proof of Proposition 7.5. □

Remark In principle, an explicit computational analysis of the minimal relations that were classified in [29, Theorem 3] should yield the optimal value of C in Proposition 7.5. In [29] it was shown that one can take $C = 7$ if one restricts p to be in the interior of the circle, and it is likely that the same bound holds in the exterior region also. However, we will not perform this computation here.

We now give a weaker version of Proposition 7.5 which is completely elementary. In particular, it avoids the Poonen–Rubinstein classification of tuples of twelve roots of unity summing to zero. It can be used as a substitute for that proposition in the proof of Theorem 1.2 (and hence Theorem 2.2), but not in the stronger Theorem 2.4.

Proposition B.2 *Let $\Pi_n \subset \mathbb{C} \cong \mathbb{R}^2$ denote the regular n -gon consisting of the n th roots of unity. Then no point other than the origin lies on more than $O(n^{5/6})$ lines joining pairs of vertices of Π_n .*

Proof The argument here will be similar to that used at the end of Sect. 7, with the roots of unity Π_n playing the role of the coset $H \oplus x$ in that analysis.

Let p be a point other than the origin. Let the vertices of Π_n be v_1, \dots, v_n in order. Suppose that the line connecting p and v_j meets the line at infinity in the point $[-\sin \pi\theta_j, \cos \pi\theta_j, 0]$. As j ranges between 1 and n , the θ_j can be taken to be an increasing sequence in $[0, 2]$.

A Euclidean geometry exercise, left as an exercise to the reader, confirms the following claim: for any fixed integer a and for any ϕ there are at most two values of j such that $\theta_{j+a} - \theta_j = \phi$.

Suppose that p lies on δn lines joining pairs of vertices of Π_n . If v_j is one such vertex, then from elementary trigonometry we see that $n\theta_j$ is an integer. Thus there is a set $J, |J| = \delta n$, such that all $n\theta_j$ are integers in $\{1, \dots, 2n\}$ for $j \in J$. Split $\{1, \dots, 2n\}$ into $m \sim \frac{1}{5}\delta n$ intervals of length $\sim 10/\delta$, and suppose that the number of points of J in these intervals is N_1, \dots, N_m . Since $N_1 + \dots + N_m = \delta n$, the Cauchy–Schwarz inequality implies that $N_1^2 + \dots + N_m^2 \geq \delta^2 n^2 / m \geq 5\delta n$. On the other hand this sum is at most the number of pairs in $J \times J$ differing by at most $10/\delta$. The contribution from the diagonal (pairs (j, j)) is just δn , and so there are at least $4\delta n$ pairs in $J \times J$ differing by at most $10/\delta$. By the pigeonhole principle there is some integer $a, 0 < a \leq 10/\delta$, such that $j, j + a \in J$ for $\gg \delta^2 n$ values of j . From

this sequences of js , we may then extract a subsequence $j_1 < \dots < j_d, d \gg \delta^3 n$, with $j_{i+1} > j_i + a$, such that once again $j_i, j_i + a \in J$ for each i .

Write $x_i := n(\theta_{j_i+a} - \theta_{j_i})$. Then, since the θ_j are increasing as a function of j , all the x_i are positive. Furthermore we have $x_1 + \dots + x_d \leq n$. However the x_i are all integers, and no integer can occur more than twice as a value of x_i by the claim we established at the start of the proof. From this it follows that $d \ll \sqrt{n}$.

Comparing these inequalities yields $\delta \ll n^{-1/6}$, and this completes the proof. \square

References

1. Bacharach, I.: Ueber den Cayley'schen Schnittpunktsatz. *Math. Ann.* **26**, 275–299 (1886)
2. Beck, J.: On the lattice property of the plane and some problems of Dirac, Motzkin, and Erdos in combinatorial geometry. *Combinatorica* **3**, 281–297 (1983)
3. Bix, R.: *Conics and Cubics: A Concrete Introduction to Algebraic Curves*. Undergraduate Texts in Mathematics, 2nd edn. Springer, New York (2006)
4. Borwein, P., Moser, W.O.J.: A survey of Sylvester's problem and its generalizations. *Aequationes Math.* **40**(2–3), 111–135 (1990)
5. Burr, S., Grünbaum, B., Sloane, N.J.: The orchard problem. *Geom. Dedicata* **2**, 397–424 (1974)
6. Carnicer, J.M., Gasca, M.: Cubic pencils of lines and bivariate interpolation. *J. Comput. Appl. Math.* **219**(2), 370–382 (2008)
7. Carnicer, J.M., Godés, C.: Generalised principal lattices and cubic pencils. *Num. Algorithms* **44**(2), 133–145 (2007)
8. Cayley, A.: *On the Intersection of Curves*. Cambridge University Press, Cambridge (1889)
9. Chasles, M.: *Traité des Sections Coniques*. Gauthier-Villars, Paris (1885)
10. Crowe, D.W., McKee, T.A.: Sylvester's problem on collinear points. *Math. Mag.* **41**, 30–34 (1968)
11. Csima, J., Sawyer, E.T.: There exist $6n/13$ ordinary points. *Discrete Comput. Geom.* **9**(2), 187–202 (1993)
12. Dirac, G.: Collinearity properties of sets of points. *Q. J. Math.* **2**, 221–227 (1951)
13. Eisenbud, D., Green, M., Harris, J.: Cayley–Bacharach theorems and conjectures. *Bull. Am. Math. Soc.* **33**, 295–324 (1996)
14. Elekes, G., Nathanson, M., Ruzsa, I.: Convexity and sumsets. *J. Number Theory* **83**(2), 194–201 (2000)
15. Erdős, P., Purdy, G.: Extremal problems in combinatorial geometry. In: *Handbook of Combinatorics*, vol. 1, pp. 809–874. Elsevier, Amsterdam (1995)
16. Fournier, J.J.F.: Sharpness in Young's inequality for convolution. *Pac. J. Math.* **72**(2), 383–397 (1977)
17. Freĭman, G.: *Foundations of a structural theory of set addition*. Translated from the Russian. *Translations of Mathematical Monographs*, vol. 37, vii+108 pp. American Mathematical Society, Providence, RI (1973)
18. Gallai, T.: Solution to problem number 4065. *Am. Math. Mon.* **51**, 169–171 (1944)
19. Gowers, W.T.: A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.* **8**(3), 529–551 (1998)
20. Hansen, S.: *Contributions to the Sylvester–Gallai theory*. Ph.D. Thesis, University of Copenhagen, Copenhagen (1981)
21. Jackson, J.: *Rational Amusement for Winter Evenings*. Longman, Hurst, Rees, Orme and Brown, London (1821)
22. Kelly, L., Moser, W.: On the number of ordinary lines determined by n points. *Can. J. Math.* **10**, 210–219 (1958)
23. Kneser, M.: Abschätzung der asymptotischen Dichte von Summenmengen. *Math. Zeit.* **58**, 459–484 (1953)
24. Kneser, M.: Ein Satz über abelschen Gruppen mit Anwendungen auf die Geometrie der Zahlen. *Math. Zeit.* **61**, 429–434 (1955)
25. Melchior, E.: Über Vielseite der projektiven Ebene. *Deutsche Math.* **5**, 461–475 (1940)
26. Motzkin, Th.: The lines and planes connecting the points of a finite set. *Trans. Am. Math. Soc.* **70**, 451–464 (1951)

27. Nilakantan, N.: Extremal problems related to the Sylvester–Gallai theorem. In: *Combinatorial and Computational Geometry*. MSRI Publications, Berkeley (2005)
28. Pach, J., Sharir, M.: *Combinatorial geometry and its algorithmic applications: the Alcalá lectures*. In: *AMS Mathematical Surveys and Monographs*, vol. 152, 235 pp. American Mathematical Society, Providence, RI (2009)
29. Poonen, B., Rubinstein, M.: The number of intersection points made by the diagonals of a regular polygon. *SIAM J. Discrete Math.* **11**(1), 135–156 (1998)
30. Silverman, J., Tate, J.: *Rational Points on Elliptic Curves*. Undergraduate Texts in Mathematics, x + 281 pp. Springer, New York (1992)
31. Sudakov, B., Szemerédi, E., Vu, V.: On a problem of Erdős and Moser. *Duke Math. J.* **129**(1), 129–154 (2005)
32. Sylvester, J.: Mathematical Question 2571. *Educational Times* (1868)
33. Sylvester, J.: Mathematical Question 11851. *Educational Times* (1893)
34. Tao, T.C.: Pappus’s theorem and elliptic curves. <http://terrytao.wordpress.com/2011/07/15/pappuss-theorem-and-elliptic-curves/>
35. Tao, T.C., Vu, V.H.: *Additive combinatorics*. In: *Cambridge Studies in Advanced Mathematics*, vol. 105, xviii+512 pp. Cambridge University Press, Cambridge (2006)