

# Gyuri Elekes and the incidences

Miklós Simonovits and Endre Szabó

April 29, 2010

Our old friend, Gyuri Elekes<sup>1</sup> was already very ill when we have finished one of *his* last paper<sup>2</sup> [3]. He died within a month. *This* paper is a continuation of [5], that contains some personal memories of the first author. Here we describe our joint work, partly published in [3] and also a basic ingredient of it, a joint paper of Elekes and Szabó [4]. The papers [3] and [4] seemed to be very important for Gyuri Elekes.

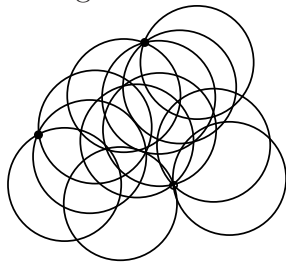
When Elekes learnt that he was soon to die, in a very short period he has finished seven papers. Two of them were these papers [4] and [3].

## 1. The beginning

In January 2006 Elekes wanted to prove Theorem 1 below and for this he needed that some functional relation of the form

$$F(\varphi_1(u, v), \varphi_2(u, v), \varphi_3(u, v)) = 0$$

cannot be satisfied (see [4, 3]) unless  $F$  has a very special form. Investigating the singularities of the functions  $\varphi_i$ , Elekes and Simonovits proved that<sup>3</sup>



**Theorem 1.** *There exist two constants  $c > 0$  and  $\eta > 0$  such that if we consider three non-collinear points in the plane,  $A$ ,  $B$ , and  $C$ , and we have through each of them  $n$  unit circles,<sup>4</sup> then the number of triple points, i.e. points belonging to three such circles, is at most  $cn^{2-\eta}$ .*

---

<sup>1</sup>More precisely, György Elekes.

<sup>2</sup>More precisely, this three-author paper was written up by Gyuri.

<sup>3</sup>A more detailed story is described in [5]

The motivation of this was the following.

Consider a square grid arrangement in the plane and  $n$  horizontal,  $n$  vertical straight lines, and  $n$  of slope 1 (i.e. at  $45^\circ$ ). They will have  $cn^2$  triple points. This can be seen in Figure 1(b). Moreover, using  $k \geq 3$  appropriate slopes we obtain  $k$  families of straight lines for which the number of crossings of multiplicity  $k$  is  $c_k n^2$ .

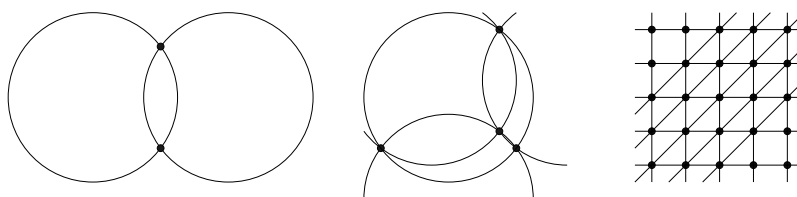


Figure 1: (a) Circles (b) and straight lines

This makes a combinatorial distinction between straight lines and *unit* circles. For arbitrary circles such a distinction cannot be made, since the straight lines can be transformed into circles by an inversion that keeps the incidences.<sup>5</sup>

Having this, we wanted to find out, when and why can such systems of curves have  $cn^2$  triple points. We wished to show that the number of triple points is small in the typical cases: if we have many triple points then we have a “strongly degenerate” case.

The proof of Theorem 1 and some of its generalizations heavily used (and also influenced?) a result of Elekes and Szabó [4]. This is why three of us decided to finish our work together. We had also another reason to join our forces: we three had three different ways of looking at the subject, and the various approaches helped each other.

We have already mentioned that Elekes worked a lot and hard in his last weeks on these two papers (too). Writing that “we have finished our paper” was a slight cheating. Actually, we have finished and published its first version [3], where we avoided using analytic branches of inverse functions of polynomials. This paper has already appeared. However, we still

<sup>4</sup>From now on, we shall write “ $n + n + n$  curves” in such cases.

<sup>5</sup>The opposite direction does not hold: two points can be contained in arbitrary many circles and this incidence pattern cannot be obtained using straight lines.

have to finish a more general, deeper version of our paper. Technically speaking, the difference between these versions is that in the first version we restrict our consideration to *explicitly parameterized* 1-parameter families of curves, while the second version will settle the problem of *implicitly parameterized* families.<sup>6</sup> Elekes felt that whenever one can apply the theorem to implicitly parameterized families of curves, then the curves can be cut into some subsegments, so the families can be replaced by explicitly parameterized families, and applying the special version (on explicitly parameterized curves) gives the same conclusion. This would mean that the practical difference between the “explicit” and “implicit” versions is not that much, at least from the point of view of applicability. The first author of this paper felt that the “implicit” version is much more natural and nicer.

The main question is as follows.

Given a 1-parameter family of “nice” curves, it may happen that they have  $cn^2$  triple points. We have seen this in case of straight lines. Applying inversion, we saw that this may happen for circles, too. We have stated that this cannot occur in case of unit circles, pierced by three distinct given points.

So the main question asks:

Given three 1-parameter families of nice curves, when can we have  $cn^2$  triple crossings and why?

**Remark 2.** The properties we investigate are – in some sense – invariant under “nice” transformations. “Nice” could mean continuously differentiable, or  $C^\infty$ , however, for us “nice” means analytic, or algebraic. If we have three nice families of curves with many triple points, we may apply any (nice) transformation<sup>7</sup>

$$u := f(x, y), \quad v := g(x, y) \tag{1}$$

to these families to get “nice” curves with the same number of triple points: If the transforming functions are nice then the new families will also be a 1-parameter families, smooth (nice) and appropriate  $n + n + n$  curves of the new families will again have  $\approx cn^2$  triple crossings.<sup>8</sup>

---

<sup>6</sup>Some definitions will follow only afterwards. Here we yield the definitions in this footnote but later we return to this. A family  $\{\gamma_t\}$  of plane curves given in the form  $F(x, y, t) = 0$  is implicitly parameterized, but if this family is parameterized in the form  $t = f(x, y)$  then it is explicitly parameterized.

<sup>7</sup>and not just the “inversion”

<sup>8</sup>Of course, we may use any – not so nice – transformations as well. However, here we are interested only in nice families.

## 2. Incidences

Elekes was interested above all, — at least in the area described here — in the following.

If we have a general family  $\mathcal{A}$  of plane curves and we select  $n$  curves from  $\mathcal{A}$ , then the number of double-points can easily be quadratic, (i.e.  $> cn^2$ ), however, the number of triple points is (mostly) relatively small. We have seen that in case of straight lines we may have  $cn^2$  triple points. However, this is a degeneracy, some kind of a strange coincidence. Which type of degeneracies should be excluded to ensure that the number of triple points be small?

Investigations of this type, on the incidence structures of straight lines and points, — or more general families of curves — belong to the fundamental questions of combinatorial geometry and perhaps the famous English mathematician, Sylvester was the first to investigate them, approximately 140 years ago. He proved that

**Theorem 3** (Sylvester). *Let  $\mathcal{L}$  be a system of  $n$  straight lines in  $\mathbb{R}^2$ . Then the number of triple points is at most  $n^2/6 + O(n)$ , and this is sharp.*

If we ask for the maximum number of triple points in case of unit circles, then not only the number of triple points but the number of double points (intersections) is at most  $n(n-1)$ . Moreover, if we restrict ourselves to curves where any two intersect in at most  $B$  points, then the number of intersections is still at most  $O(n^2)$ . Answering a question of Paul Erdős, Elekes proved that

**Theorem 4** (Elekes [2]). *There exist  $n$  unit circles with at least  $cn^{3/2}$  triple points.*

The upper and lower bounds are rather far from each other.

We could conjecture that if we consider “nice” families of curves that are not straight lines, then we may have only  $o(n^2)$  triple points. This is not so: the family of straight lines can be transformed into families of *congruent parabolas* with  $cn^2$  triple crossings. (Or — as we have mentioned — it can be transformed into families of circles.)

On the other hand, if we consider nice families of curves and exclude certain degeneracies, then we can prove that the number of triple points is only at most  $cn^{2-\eta}$ , for some suitable constants  $c > 0$  and  $\eta > 0$ .

Here we shall explain our results, and the background, without trying to formulate them in their most general form. We should clarify – among others – two questions:

- Which are the nice curves and nice families of curves?
- Which degenerate cases should be excluded to get only  $o(n^2)$  triple crossings.

## 2.1. Nice families of curves and their Enveloping curves

Below we wish to explain the most important phenomena. Therefore occasionally our formulations will be slightly loose, often heuristic. In our next paper, on the implicitly parameterized families, we shall take the effort to be 100% precise.

We start with defining the curves, families of curves and the main content of our results. Here a *curve*  $\gamma$  is always the 0-set of some polynomial  $P$  of two variables, in the plane:

$$\gamma = \left\{ (x, y) \in \mathbb{R}^2 \mid P(x, y) = 0 \right\} \quad (2)$$

Of course, the questions we consider can be considered above any field, primarily above the reals and complex numbers. However, we assume that most of our readers prefer the real curves. Thus here we shall restrict ourselves (with two exceptions) to the real case. The curves will be denoted by script capital letters:  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$ ,  $\mathcal{E}$ , or sometimes by  $\gamma$ .

If we start changing smoothly some parameter in the equation (2) describing a curve, then the curve itself often starts changing smoothly. This way we get a family of curves. If the parameter is  $t$ , then we mostly consider families of curves described by polynomials depending on  $t$ :

$$\gamma_t = \left\{ (x, y) \in \mathbb{R}^2 \mid P_t(x, y) = 0 \right\} \quad (3)$$

Here we are interested only in families of curves where the coefficients are polynomials of a parameter  $t$ :  $P_t(x, y) = \tilde{P}(x, y, t)$ , where  $\tilde{P}$  is a polynomial of the three variables. The families usually will be denoted by script capital letters, like  $\{\mathcal{A}_t\}$ ,  $\{\mathcal{B}_t\}$ , where the index is the parameter: If  $t_1 \in \mathbb{R}$  is a parameter value, then  $\mathcal{A}_{t_1}$  denotes the member of the family  $\{\mathcal{A}_t\}$  corresponding to  $t = t_1$ .

**Example 5.** Whatever we write here is valid for arbitrary families of curves. However, it is worth keeping in mind the following case that is slightly more

general than the unit circles in Theorem 1. We are given a point  $A$  in the plane and a convex closed curve  $\mathcal{A}_0$ , defined by a polynomial. We rotate  $\mathcal{A}_0$  around  $A$ , getting a family of curves, see Figure 2. We do not assume  $A \in \mathcal{A}_0$ . Denote by  $\mathcal{A}_t$  the curve rotated by  $\alpha$ , for  $t = \tan \frac{\alpha}{2}$ .

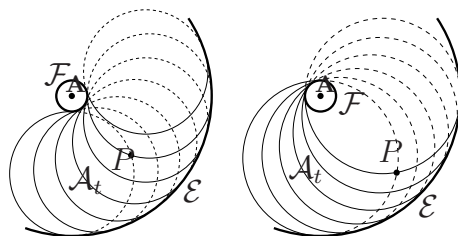


Figure 2: Families of convex curves, two situations

The reader could ask, why did we use  $t = \tan \frac{\alpha}{2}$ , to parametrize the curves, instead of using  $\alpha$  itself. The answer is that in the equations of the rotated curves we have  $\sin \alpha$  and  $\cos \alpha$ . They are not polynomials of  $\alpha$ . However, if we use  $t = \tan \frac{\alpha}{2}$ , then we shall obtain polynomials of  $t$ , after some work. Here we emphasize that the actual form of these polynomials is completely irrelevant for us, we need only the existence of these polynomials. Moreover, the proofs will neither use these polynomials.

The Reader can see two situations in Figure 2: On the right the center  $A$  (of the rotation) is inside of  $\mathcal{A}_t$ , on the left it is outside. The curves cover a ring type domain in both cases. If  $A$  happened to be on  $\mathcal{A}_0$ , then the inside circle of this ring would shrink to a point, so the rotated curves would cover a disk.

One can see two special curves on this figure:  $\mathcal{E}$  and  $\mathcal{F}$ . They are the enveloping curves of these two families. In general,

**Definition 6.** We call a curve  $\mathcal{E}$  the *enveloping curve* of  $\{\mathcal{A}_t\}$  if in each point of  $\mathcal{E}$  some curve  $\mathcal{A}_t$  of the family is tangent to  $\mathcal{E}$ , however, no  $\mathcal{A}_t$  has an arc common with  $\mathcal{E}$ .

So nearby points of  $\mathcal{E}$  are tangent-points of distinct curves  $\mathcal{A}_t$ , hence infinitely many curves  $\mathcal{A}_t$  are tangent with  $\mathcal{E}$ .<sup>9 10</sup>

<sup>9</sup>Observe that we have formulated our definition so that a non-degenerate sub-arc of an enveloping curve is also an enveloping curve, where degenerate means a point or an empty arc.

<sup>10</sup>If the parameterized family is given in form of  $F(x, y, t) = 0$ , then the enveloping

We have subdivided each curve  $\mathcal{A}_t$  of Figure 2 into two arcs, one indicated by a thinner line the other by a thicker but broken line. Each point in the domain covered by the whole curves is covered by two such half-curves: a thin continuous line and a thick broken line.

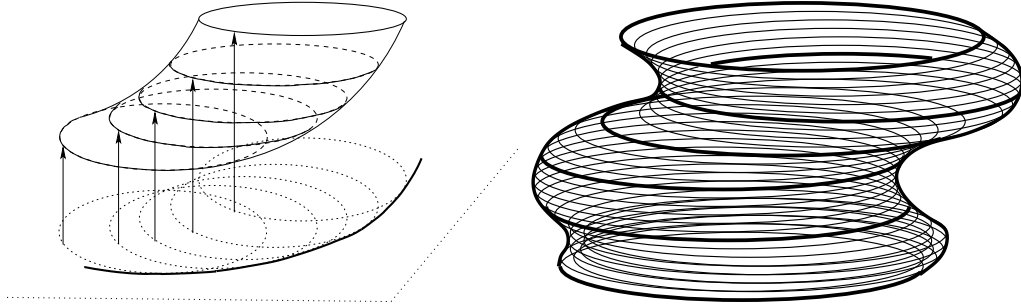


Figure 3: The lifted curves define a “spiral column”

It would simplify the situation if we could achieve that the points on a curve determine the parameter of that curve: each point is only on (at most) one curve  $\mathcal{A}_t$ . Staying in  $\mathbb{R}^2$ , we cannot achieve this. Therefore we apply a “lifting”: the curve  $A_t$  (see Figure 3) is lifted from the plane into the space, into height  $t$ . These space-curves constitute a spiral tube (similar to the barock columns). Let

$$V = \left\{ (P, t) \in \mathbb{R}^2 \times \mathbb{R} \mid P \in \mathcal{A}_t \right\} .$$

Clearly, if we draw our curves onto the surface  $V$ , then they become non-intersecting: each point of the surface belongs only to one curve. Let us return to the problem of triple points.

**Definition 7.** Given three families of curves:  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$ , and  $\{\mathcal{C}_t\}$ , we shall say that they are “**in special position**”, if for some constant  $c > 0$ , for infinitely many integers  $n$  we can find three  $n$ -tuples of parameters,  $X, Y, Z \subset \mathbb{R}$  for which “the generalized grid”  $X \times Y \times Z \subset \mathbb{R}^3$  has at

---

curves are (basically) described by

$$F(x, u, t) = 0 \quad \text{and} \quad \frac{\partial}{\partial t} F(x, y, t) = 0 .$$

We have to remark here that there are several ways to define the enveloping curves and they are not completely equivalent.

least  $cn^2$  parameter-triples  $(r, s, t)$  (out of the  $n^3$  possible ones) that are triple points:

$$\mathcal{A}_r \cap \mathcal{B}_s \cap \mathcal{C}_t \neq \emptyset.$$

Actually, a point can correspond to many triple intersections and we are counting the points, not the parameter triples.

If our curves are not “in special position”, then we say that they are “**in typical position**”.

These definitions are motivated by that mostly (?) the number of triple points is small. Thus, e.g., the families of circles in Theorem 1 are “in typical position”. Observe this fairly surprising jump in the number of triple points: three such families of curves either have (at least)  $cn^2$  triple points, for suitably chosen parameters, or the families are “in typical position”, and then — whichever way we choose the three  $n$ -tuples  $X, Y$ , and  $Z$  above, — the generalized grid  $X \times Y \times Z \subset \mathbb{R}^3$  will give at most  $cn^{2-\eta}$  parameter-triples  $(r, s, t)$  corresponding to triple intersections.<sup>11</sup> We shall see both behaviours below. In Example 8 we shall see families of parallel straight lines “in special position”. On the other hand, in Example 9 we see families of circles “in typical position”. This will lead us to the generalization of Theorem 1.

**Example 8.** Let  $\mathcal{A}_r, \mathcal{B}_s, \mathcal{C}_t$  denote the horizontal line  $y = r$ , the vertical line  $x = s$  and the skew line  $y = x + t$ , respectively. They form three families:  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$  and  $\{\mathcal{C}_t\}$ . In Figure 1(b) one can see this three families of lines, and it is easy to count the triple-crossings. It is clearly a “special arrangement of curves”: for example if  $X = Y = Z = \{a, 2a, 3a, \dots, na\}$  for an arbitrary value  $a > 0$ , then in the “generalized grid”  $X \times Y \times Z$  there are  $cn^2$  triple-crossings.

These three families of lines will play an important role. We shall see in Theorem 16 that each “special arrangement of curves” can be reconstructed<sup>12</sup> via clever transformations from this single configuration.

**Example 9.** In Figure 4(a) one can see three copies of the families of curves of Example 5. Given three centers of rotation and three closed convex curves, we rotate each of the three curves around the center assigned to it. This way we get three families of curves. Denote them by  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$  and  $\{\mathcal{C}_t\}$ .

<sup>11</sup>If  $n$  is large, then  $n^2$  is much larger than  $n^{2-\eta}$ . So either we have very many triple points, or very few: there are no in-between situations. Similar phenomena occur in some other combinatorial situations as well.

<sup>12</sup>Basically it contains an image of such a configuration



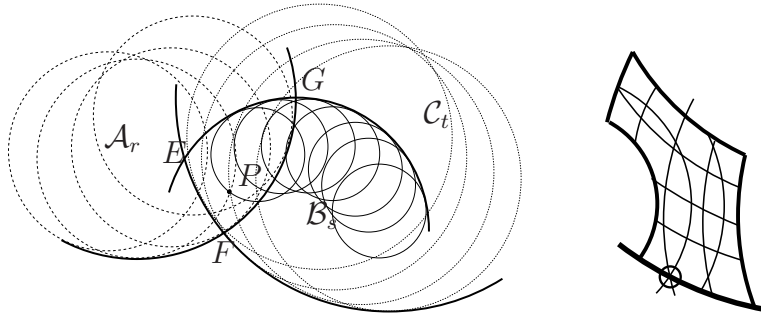


Figure 4: (a) Three families of circles (b) circular pentagon

In Figure 4(a) the letter  $P$  marks a triple-crossing of three curves, one from each family: one curve is drawn continuously, the other one is dashed and the third one is dotted. Each of the three families sweeps through an annulus, the domain of triple-crossing points is the intersection of the three annuli. On the left it is just the “curvy triangle”  $EFG$ . In general, the situation can be more complicated. E.g., if we slightly shrink  $\mathcal{B}_0$ , the family  $\{\mathcal{B}_s\}$  will not reach  $F$ . Hence the triple-crossing points will not cover this corner of the domain  $EFG$ . There are cases where the domain of triple-crossing points is even more complicated, see Figure 4(b).

**Theorem 10.** *Suppose that in Example 9 the three families of curves are arranged according to Figure 4(a), i.e. the set of triple-crossings is precisely the “curvy triangle”  $EFG$ . Then our configuration is a “typical arrangement of curves”, hence  $n - n - n$  curves can have at most  $cn^{2-\eta}$  triple-crossings for some suitable constants  $c > 0$  and  $\eta > 0$ .*

Naturally there are many other “special arrangements of curves” besides the one in Example 8. We have seen in Remark 2, how to get, with the help of continuous transformations, numerous new “special arrangements of curves” from a single one. Presently we use only transformations that can be given via polynomial functions, and their image set is two dimensional. (This later condition excludes, e.g., the projection to a line.) They are called *polynomial transformations*.

**Example 11.** Let  $f(x, y) = x^2$  and  $g(x, y) = y^3 - 3y$  in (1) in Remark 2. This is an extremely special situation, since the transforming functions depend only on one variable: we may study separately the horizontal and vertical behaviour of this transformation. On the one hand, the function  $x^2$  maps  $\mathbb{R}$  into itself “folding” the negative side back onto the positive side.

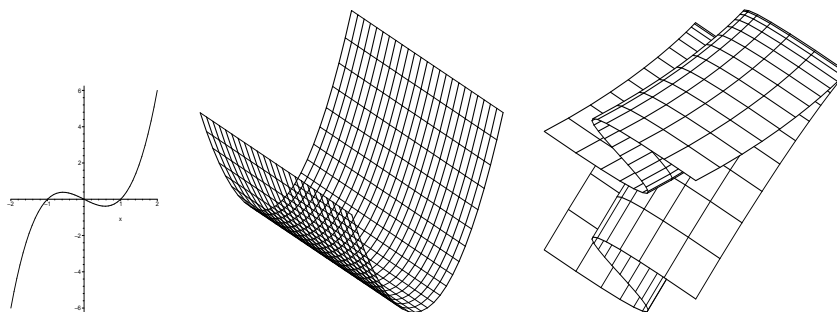


Figure 5: Folding the plane

Therefore our transformation also “folds the plane in half” along the vertical line  $x = 0$ . On the other hand, the function  $y^3 - 3y$  is increasing on the half-line  $(1, \infty)$ , then “turns back” and decreases on the interval  $(-1, 1)$ , then “turns” again and increases on the remaining half-line  $(1, \infty)$ . Therefore our transformation also “folds” the plane twice in the  $y$  direction: along the horizontal lines  $y = -1$  and  $y = 1$ . One can imagine a sheet first folded in half along a vertical line, then “wrinkled” along two (nearby) horizontal lines, and finally “ironed” into the plane. Our transformation works in the plane in much the same way, but there is a subtle difference. If one cuts the sheet into rectangles along the fold-lines, then the ironing smoothes each rectangle congruently into the plane. However, our transformation “distorts” the rectangles both horizontally and vertically: some parts get stretched out, others are shrunk.

This example illustrates well the behaviour of other polynomial transformations — the most important difference being that the fold-lines are no longer straight lines. Let  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be any polynomial transformation. The locus of critical points, where the Jacobian determinant of  $\phi$  vanishes, consists of finitely many curve components and finitely many isolated points. The curve part is the correct generalization of the fold-lines of Example 11,  $\phi$  will actually “fold” the plane along some of these curves. One can add a few more curves to the picture so that these curves together “cut up” the plane into “curvy polygonal domains”, and  $\phi$  is one-to-one on each of these domains. The addition of extra curves is necessary as shown in the example  $(x, y) \rightarrow (x^2 - y^2, 2xy)$ , which is simply the map  $z \rightarrow z^2$  of the complex plane, and has a single critical point at the origin.

Univariate polynomials usually have no inverse function, hence we often use “multi-valued inverse functions”. Good examples are the formulas

for solving quadric, cubic or quartic equations which are two, three and four-valued. More generally, the inverse of a polynomial of degree  $d$  is (at most)  $d$ -valued. Analogously, we can invert polynomial transformations. Their inverses are “multiple-valued transformations” which send each point to multiple images<sup>13</sup>. One can easily estimate the number of images. There can be finitely many exceptional points in the plane whose image is an entire curve (hence infinitely many points), but all other points can have at most  $\deg(f) \cdot \deg(g)$  images,<sup>14</sup> where  $f$  and  $g$  denote the polynomials defining the transformation to be inverted (see formula (1) of Remark 2).

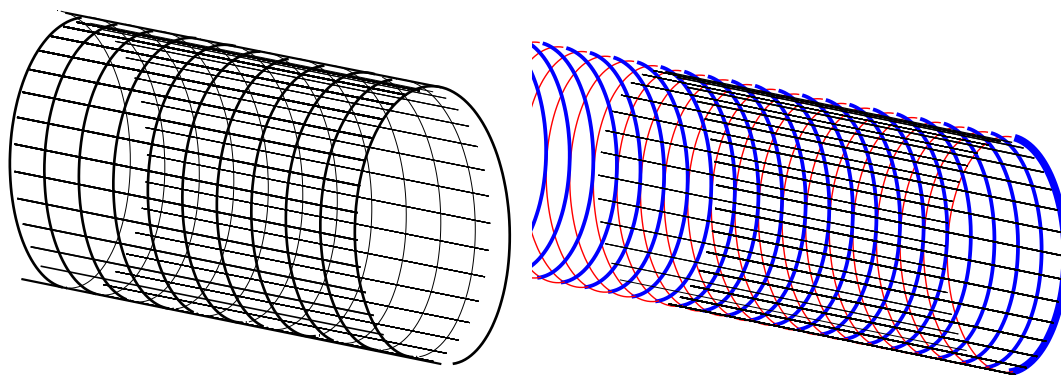


Figure 6: (a) Spiral and “parallel lines” (b) Grid on the cylinder

There is another possibility that we haven’t used so far: one can draw not only in the plane, but on a sphere, on a cylinder, or on any other smooth enough surface. We met such a drawing, e.g., in Figure 3. In this paper we use *smooth algebraic surfaces*, that is, surfaces defined via polynomial equations which have a tangent plane at each of their points.

**Example 12.** Let us consider again Example 8 and the corresponding Figure 1(b). If we “roll up” this drawing vertically then we get a horizontally lying cylinder and three families of curves drawn on it: The horizontal lines turn into a horizontal ruling of the cylinder, the vertical lines turn into vertical circles (the orthogonal cross-sections of the cylinder), and the skew

<sup>13</sup>More precisely: a single point can have zero, one, or many images.

<sup>14</sup>A system of two bivariate polynomial equations have either infinitely many solution or at most as many as the product of the degrees of the equations. (Bezout Theorem?)

lines become spirals running round-and-round. They together form another “special arrangement of curves”.

If we “roll up” the cylinder again around a vertical axis, then we get a *torus*. The horizontal lines turn into horizontal circles, the vertical circles remain vertical circles (but they are no longer parallel, they revolve around the new axis) and the spirals on the cylinder will turn into spirals on the torus. This is again a “special arrangement of curves”.

The spirals drawn onto the torus can close into themselves, or can run round-and-round the torus indefinitely — it depends on the relationship between the two “rolling ups”. For us only the self-closing spirals are important, since they can be defined with polynomial equations. (The actual form of these polynomials is, as usual, irrelevant.)

It is worth contemplating for a moment: if a spiral drawn on a torus does not close into itself, then it travels all over the torus uniformly, and its image is a dense subset. These dense spirals are impossible to define with polynomial equations. Indeed, if a polynomial vanishes along a dense subset, then it is zero on the entire torus.

### 3. Reasons to be special

**Definition 13.** We shall use the name *line-like arrangement of curves* for the three kinds of curve configurations we met in Examples 8 and 12. Let us note that none of the families in these arrangements has enveloping curve.

Soon we shall see that all “special arrangement of curves” can be obtained from one of the line-like arrangements of curves via polynomial transformations and their inverses: this is the meaning of Theorem 16. As a matter of fact, we constructed each line-like arrangement of curves via a continuous transformation from the line families of Example 8, hence all “special arrangement of curves” originates eventually from this single one. But the transformation used in Example 12, the “rolling up”, is not a polynomial transformation. (A periodic function cannot be a polynomial.) In this paper we prefer to use polynomials only, this is why we need three “basic” arrangements instead of just one.

**Example 14.** Given a line-like arrangement of curves on a surface  $F$ . (Hence  $F$  is either the plane, or a cylinder, or a torus.) We would like to build from it as many “special arrangements of curves” as we can. Let us

choose an arbitrary smooth algebraic surface  $W$  and two polynomial transformations:  $\phi : W \rightarrow \mathbb{R}^2$  and  $\psi : W \rightarrow F$ . Let us denote by  $\phi(\psi^{-1}(\_))$  that “multiple-valued transformation” that we get by first applying the inverse of  $\psi$  (this is multiple valued, denoted by  $\psi^{-1}$ ), and then applying  $\phi$ . This composed transformation maps the surface  $F$  first into  $W$  and then transforms it further into the plane. Naturally, the line-like arrangement of curves drawn on  $F$  gets also transformed into a “special arrangement of curves” in the plane.

In an ideal world one would hope that the main reason for a configuration to be special is that it contains some kind of image of the three line families of Example 8. This is essentially true, as we shall see, apart from a few exceptions.

Even if a configuration is a “special arrangement of curves”, we cannot expect that all of its portions be “nice”. It may happen that someone adds a few “unnecessary” arcs<sup>15</sup> to an already nice configuration — this will certainly not decrease the number of triple-crossings. Therefore the most we can hope for is that from each “special arrangement of curves” one can select the “essential” arcs, and this essential part behaves already “truly well”.

**Definition 15.** Let  $\{\mathcal{A}_r\}$  and  $\{\mathcal{B}_s\}$  be two families of plane curves. We say that *they have a common component*, if for all values of  $r$  there is a value of  $s$  such that the curves  $\mathcal{A}_r$  and  $\mathcal{B}_s$  have a common arc.<sup>16</sup>

**Theorem 16.** *Suppose that the families  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$ ,  $\{\mathcal{C}_t\}$  form a “special arrangement of curves”. Then, as in the Example 14, there exists*

- a line-like arrangement of curves  $\{\overline{\mathcal{A}_{\bar{r}}}\}$ ,  $\{\overline{\mathcal{B}_{\bar{s}}}\}$  and  $\{\overline{\mathcal{C}_{\bar{t}}}\}$  drawn on a surface  $F$  (so  $F$  is either the plane, or a cylinder, or a torus),<sup>17</sup>
- a smooth algebraic surface  $W$  and two polynomial transformations  $\phi : W \rightarrow \mathbb{R}^2$  and  $\psi : W \rightarrow F$

*such that the “multiple-valued transformation”  $\phi(\psi^{-1}(\_))$  sends the family  $\{\overline{\mathcal{A}_{\bar{r}}}\}$  to a family of plane curves that have a common component with the*

<sup>15</sup>The extra arcs are sometimes the continuations of the original arcs, but one can add brand new components as well.

<sup>16</sup>The definition suggests that this relation is symmetric, however, one has to prove this, under some additional conditions.

<sup>17</sup>We use the notation  $\bar{r}$ ,  $\bar{s}$  and  $\bar{t}$  for these parameters to distinguish them from the original  $r$ ,  $s$  and  $t$  parameters. The theorem states that each curve  $\mathcal{A}_r$  is related in a certain way to some other curve  $\overline{\mathcal{A}_{\bar{r}}}$ , but it does not tell anything about the relationship between the two parameter values  $r$  and  $\bar{r}$ .

family  $\{\mathcal{A}_r\}$ , and similarly, the transformed image of  $\{\overline{\mathcal{B}}_s\}$  and  $\{\overline{\mathcal{C}}_t\}$  have a common component with  $\{\mathcal{B}_s\}$  and  $\{\mathcal{C}_t\}$  respectively. (One can choose the arcs of these common components in such a way that they too form a “special arrangement of curves”.)

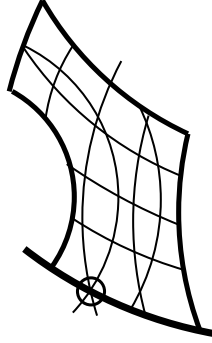
We shall close this paper with a variant of this theorem, (namely Theorem 20) as it was used in [3]. Here we shall continue with some topological explanations. Much of this – as we stated – is on the level of heuristic arguments that can be turned into precise arguments.

## 4. Reasons to be typical

It turns out that there is a very general principle, a good geometric explanation for the “typical” behaviour of most curve configurations. On the one hand, a family of plane curves usually has plenty of envelopes. On the other hand, envelopes are scarce inside a “special arrangement of curves”. We shall use an irreducibility condition to formulate this principle precisely. A priori an enveloping curve can have information only about those arcs in the family members which are tangent to it — but the irreducibility will ensure us that this information automatically “spreads all over”: in an irreducible family almost all the arcs behave uniformly.

**Definition 17** (Concurrency). Given three families of curves in the plane. We say that “**the graph of triple-crossings is irreducible**”, if there is an irreducible trivariate polynomial  $\Psi(r, s, t)$  that vanishes in all the parameter-points  $(r, s, t)$  corresponding to the triple-crossings.  $\Psi$  is called the **concurrency function**.

At the moment, this definition may seem too algebraic to be useful. To remedy the situation, we shall describe a simple geometric test for irreducibility. The spatial surface given by the equation  $\Psi(r, s, t) = 0$  could be called the graph of triple-crossings, but here we use slightly different language instead.



**Theorem 18.** *Given three families of curves:  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$  and  $\{\mathcal{C}_t\}$ . Suppose that one of them, say  $\{\mathcal{C}_t\}$  has an envelope  $\mathcal{E}$ , and this envelope meets in a single point two members, say  $\mathcal{A}_{r_0}$  and  $\mathcal{B}_{s_0}$  from both of the other families (here  $r_0$  and  $s_0$  denote appropriate parameter values). If no two of the three curves  $\mathcal{E}$ ,  $\mathcal{A}_{r_0}$  and  $\mathcal{B}_{s_0}$  are tangent to each other, and the graph of triple-crossings is irreducible, then the three families form a “typical arrangement of curves”, i.e. three times  $n$  curves can have at most  $cn^{2-\eta}$  triple-crossings.*

This theorem, again, has several forms and variants. Below we formulate a variant of this theorem, as stated in [3] (without much explanation, in its complex version). Denote  $\text{cl}(X)$  the closure of  $X$ .

**Theorem 19** (Main Theorem). *Let  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$  and  $\{\mathcal{C}_t\}$  be families explicitly parameterized by the functions  $f_1, f_2, f_3$ , analytic<sup>18</sup> on open domains  $G_1, G_2, G_3$  and continuous on the closed domains  $\text{cl}(G_1), \text{cl}(G_2), \text{cl}(G_3)$ , respectively. Assume that  $\mathbf{G} = G_1 \cap G_2 \cap G_3$  is connected. Assume that any two curves intersect in at most  $B$  points, and the concurrency of three curves is described by a polynomial  $\Psi$  (see Definition 17). Moreover, assume that*

- (a)  $\{\mathcal{C}_t\}$  has an envelope  $\mathcal{E}$ ;
- (b)  $\mathcal{E} \subseteq G_1 \cap G_2$ ;
- (c) No  $f_i$  ( $i = 1, 2, 3$ ) is constant on any non-empty open sub-arc of  $\mathcal{E}$ .<sup>19</sup>

*Then the number of triple points is at most  $B \cdot n^{2-\eta}$ , for a suitable  $\eta = \eta(\deg(\Psi))$  — provided that  $n > n_0 = n_0(\deg(\Psi))$ .*

Let us return to Theorem 10. It is a special case of Theorem 18. We shall see later that in this situation the graph of triple-crossings is irreducible. In Figure 4 we can see three envelopes: the fattened arcs  $EF$ ,  $FG$  and  $GH$ . Let us consider, say, the arc  $EG$ . One can see well in the drawing that both the dashed curves (the members of the family  $\{\mathcal{A}_r\}$ ) and the dotted curves (the members of the family  $\{\mathcal{B}_s\}$ ) intersect the arc  $EG$ , but none of them are tangent to it. Moreover, those dashed and dotted curves which meet each other just on the arc  $EG$ , cross each-other transversally. Therefore one can apply Theorem 18 to this enveloping arc  $EG$ , proving Theorem 10. (Needless to say that one could apply it to the other two envelopes as well.)

<sup>18</sup>Here we could write polynomial instead of analytic, since we promised to restrict ourselves to polynomially defined curves.

<sup>19</sup>Intuitively: no non-empty open sub-arc of  $\mathcal{E}$  is contained in any of the considered curves of  $\{\mathcal{A}_r\}$ ,  $\{\mathcal{B}_s\}$  and  $\{\mathcal{C}_t\}$ .

#### 4.1. Some remarks about the proof of Theorem 18

Suppose, to the contrary, that the three families form a “special arrangement of curves”. According to Theorem 16, there is a “multiple-valued transformation”  $\phi(\psi^{-1}(\_))$  which rebuilds<sup>20</sup> our three families from a line-like arrangement of curves. (A pleasant consequence of irreducibility is that the transformation rebuilds our entire configuration, not just a small portion of it.) Of course, the line-like arrangement of curves contains no envelopes at all, hence our envelope  $\mathcal{E}$  “was born” during the transformation. We have discussed on page 10 after Example 11 that a polynomial transformation behaves very nicely in the complement of a certain curve  $F$  (which contains the critical curves, among them all the fold-curves, and may have further components). Each connected component  $K$  of the complement of  $F$  is mapped one-to-one into the plane. If a family of curves has no envelopes inside  $K$ , then after transformation, this portion of the family looks “essentially the same”, cannot “develop” an envelope. The same description remains valid for “multiple-valued transformations”. Therefore envelopes can “be born” only along the curve  $F$  corresponding to  $\phi(\psi^{-1}(\_))$ . Let us pretend for a moment that  $F$  consists of fold-curves only. The image of a fold-curve should be an envelope for all three families, as all curves “turn back” at the fold-curve (in fact the entire “sheet of paper” turns back there). But this possibility is excluded by the other condition in Theorem 18: the curves  $\mathcal{A}_{r_0}$  and  $\mathcal{B}_{s_0}$  cross transversally our envelope  $\mathcal{E}$ .<sup>21</sup> We ran into a contradiction, hence the three families do not form a “special arrangement of curves”. In general,  $F$  may have components which are not fold-curves. It turns out that only those components can produce envelopes which are critical (i.e. where the Jacobian of the transformation is not invertible, the fold-curves are among them), and a slight modification of the above argument applies to all such components.<sup>22</sup>

Next we introduce a very general geometric technique, the construction of ramified coverings. We shall use it to show (as promised) that the graph of triple-crossings is irreducible in Theorem 10. General techniques like this one have many applications. For example, they play a central role in the proof of Theorem 16 (that we cannot reproduce here because of its length).

---

<sup>20</sup>Rebuilds? This heuristic description means that one of the configurations of three families of lines is “embedded” into our families and therefore provides the many triple points, proving that our families are “in special position”, as explained in Examples 8, 12 and Theorem 16.

<sup>21</sup>Here the transversality of one of them would be enough, but later we need both.

<sup>22</sup>This is the point where we need the transversality of both curves  $\mathcal{A}_{r_0}$  and  $\mathcal{B}_{s_0}$ .



Let us return to Figure 4 and the families of Theorem 10. According to our assumptions, all triple-crossings live in the “curvy triangle”  $EFG$ . Hence we shall restrict our attention to this domain only and to the portion of the arcs running within that. In Figure 4 we see three arcs passing through the point  $P$ , but the drawing is not complete: each of the three families has two arcs (altogether six arcs) passing through  $P$  and through any other point of the “curvy triangle”  $EFG$ . As a contrast, in Figure 1(b), at each point there is only a single line passing through from each of the three families (altogether three lines). With our ramified coverings, we shall try to eliminate this striking difference.

We have seen in Figure 3 how “lifting” helps to separate in the space the members of a family of plane curves. Let us repeat the same trickery, say, with the family  $\{\mathcal{B}_s\}$ . Originally, in Figure 3 we have “unfolded” a planar domain (an annulus) along one of its boundary arcs (the darkened one) and obtained a domain on a spatial surface (a tube-like thing in that case) which covers doubly the original domain, and the new family drawn onto the new surface has only one curve passing through any of the points. From the whole annulus right now we are interested in the “curvy triangle”  $EFG$  only (see that in Figure 4). Therefore, instead of the whole darkened fold-curve of Figure 3, only the arc  $EG$  will play a role now. (The rest of that picture, though remains valid, is simply ignored.)



Figure 7: Folding up a flattened ball

Similarly, now we unfold (i.e. double) our “curvy triangle” along the arc  $EG$ , and the two copies smooth out into a spherical slice. This unfolding, among other things, is illustrated on the last two images of Figure 7. We have photographed the three-step folding up of a flattened ball. There are four phases in Figure 7, with a single fold taking place between any two consecutive images. In the last step we have folded in half a spherical slice, and obtained a spherical triangle — quite a bit thickened, since the sheets of a real-world ball do not usually squeeze tightly together. The final spherical triangle is very much like our “curvy triangle”  $EFG$ , so our unfolding should look much the same as the unfolding of the folded-up ball (doing backward the last step of Figure 7). One can clearly see on the photo that, after

doubling, the two copies of the front side  $EF$  of the spherical triangle smooth together into the circular arc (the front edge on the third photo), and the same thing happens with the right hand side  $FG$  (which turns into the back side edge on the third photo).

This trick separates only the curves of the family  $\{\mathcal{B}_s\}$ , they will disjointly rule the ball-slice on the third photo (similarly to Figure 3). Imagine now that we copy also the other two families of curves,  $\{\mathcal{A}_r\}$  and  $\{\mathcal{C}_t\}$  on our folded-up ball (fourth photo), the same way as they can be seen in the “curvy triangle”  $EFG$  in Figure 4. Press the pencil hard to make the curves appear on all the eight sheets. After the unfolding, each arc appears twice on the ball-slice (on the third photo), but the two copies of the arcs in the families  $\{\mathcal{A}_r\}$  and  $\{\mathcal{C}_t\}$  do not smooth out, they stay separate. Our goal is not completely achieved yet: through each point of the ball-slice there are still two arcs passing through from both the families  $\{\mathcal{A}_r\}$  and  $\{\mathcal{C}_t\}$ .

We repeat the process two more times, we “unfold completely” the ball as it is shown on the series of images (from right to left) in Figure 7. In the first step, as we have already discussed it, we have separated the arcs of the family  $\{\mathcal{B}_s\}$ . It is easy to check that with the second step we separate the arcs of the family  $\{\mathcal{A}_r\}$ , and the last step separates the arcs of the family  $\{\mathcal{C}_t\}$  as well. Hence on the fully unfolded ball, through each point there is only one arc passing through from each of the three families.

One can easily verify, that along the way each arc is quadrupled — since they are not doubled in that step when we “separated” them from their own family mates. The unfolded arcs are not necessarily smooth, they can have angles along the fold-curves.

Why did we work so hard? Of course, we had no chance to get the same arrangement as the one in Figure 1(b): after all, our configuration is not supposed to be a “special arrangement of curves”. But we had quite a different goal in mind. Within our new configuration drawn on the surface of the ball, any triple-crossing can be moved continuously into any other (the sphere is a connected surface). In fact, as we shall explain it soon, this indicates that the graph of triple-crossings is irreducible.

In general, we can play our unfolding game on any configuration along each enveloping arc. If we duly go through the whole process, finally we arrive to a surface  $W$  which is “ruled only one-fold” by each of the three families of curves. Of course, there is no guarantee that we get a sphere again. It may very well happen that our new  $W$  consists of, say, two spheres and a torus. It turns out, that if  $W$  has only a single component, then the graph of triple-crossings is irreducible. It is a much more involved task to

decide the shape of the components of  $W$ , but luckily we are not concerned now with that problem.

We have completed this unfolding for the configuration of Theorem 10 and we have obtained a single component, a sphere. Hence the graph of triple-crossings is irreducible, one can apply Theorem 18. In fact, this is the standard way to apply our methods. We always unfold our surface and study the unfolded  $W$  instead. Finally, let us reveal one more secret of the trade: even if the irreducibility condition fails, and one cannot apply Theorem 18 directly, one can still study the components of  $W$  separately and try to find for each of them an appropriate enveloping curve. With this method, most of the time it is possible to decide whether one has a “special arrangement of curves” or not.

## 5. Appendix: The Surface theorem

With this we have finished the mathematical discussions. As an “Appendix” we include that very version of Theorem which we used in [3], from [4]. Again, we skip most of the explanation. The meaning of the next theorem is that if we have many triple points then the surface  $\Psi(r, s, t) = 0$  goes through many “generalized grid points”  $(r, s, t) \in X \times Y \times Z$  and this can happen only if  $\Psi$  is basically  $r + s + t$ , apart from some coordinatewise transformation.

**Theorem 20** (“Surface Theorem”, see [4], Theorem 3.). *For any positive integer  $d$  there exist positive constants  $\eta = \eta(d) \in (0, 1)$  and  $n_0 = n_0(d)$  with the following property.*

*If  $V \subset \mathbb{C}^3$  is an algebraic surface (i.e. each component is two dimensional) of degree  $\leq d$  then the following are equivalent:*

- (a) *For at least one  $n > n_0(d)$  there exist  $X, Y, Z \subset \mathbb{C}$  such that  $|X| = |Y| = |Z| = n$  and*

$$|V \cap (X \times Y \times Z)| \geq n^{2-\eta};$$

- (b) *Let  $\mathbb{D} \subset \mathbb{C}$  denote the open unit disc. Then either  $V$  contains a cylinder over a curve  $F(x, y) = 0$  or  $F(x, z) = 0$  or  $F(y, z) = 0$  or, otherwise, there are one-to-one analytic functions  $g_1, g_2, g_3 : \mathbb{D} \rightarrow \mathbb{C}$  with analytic inverses such that  $V$  contains the  $g_1 \times g_2 \times g_3$ -image of a part of the plane  $x + y + z = 0$  near the origin:*

$$V \supseteq \left\{ \left( g_1(x), g_2(y), g_3(z) \right) \in \mathbb{C}^3 : x, y, z \in \mathbb{D}, x + y + z = 0 \right\}.$$

- (c) For all positive integers  $n$  there exist  $X, Y, Z \subset \mathbb{C}$  such that  $|X| = |Y| = |Z| = n$  and  $|V \cap (X \times Y \times Z)| \geq (n-2)^2/8$ .
- (d) Both (b) and (c) can be localized in the following sense. There is a finite subset  $H \subset \mathbb{C}$  and an irreducible component  $V_0 \subseteq V$  such that whenever  $P \in V_0$  is a point whose coordinates are not in  $H$  and  $U \subseteq \mathbb{C}^3$  is any neighborhood of  $P$ , then one may require that  $(g_1(0), g_2(0), g_3(0)) = P$  in (b), and the Cartesian product  $X \times Y \times Z$  in (c) lies entirely inside  $U$ . Furthermore,  $P$  has a neighborhood  $U'$  such that each irreducible component  $W$  of the analytic set  $V_0 \cap U'$ , with appropriate  $g_1, g_2$  and  $g_3$ , can be written in the form

$$W = \left\{ \left( g_1(x), g_2(y), g_3(z) \right) \in \mathbb{C}^3 : x, y, z \in \mathbb{D}, x + y + z = 0 \right\}.$$

If  $V \subset \mathbb{R}^3$  then the equivalence of (a), (b), (c) and (d) still holds with real analytic functions  $g_1, g_2, g_3$  defined on the interval  $(-1, 1)$ .

## References

- [1] Gy. Elekes: A geometric inequality and the complexity of computing volume. *Discrete Comput. Geom.* 1 (1986), no. 4, 289–292.
- [2] Gy. Elekes:  $n$  points in the plane can determine  $n^{3/2}$  unit circles. *Combinatorica* 4 (1984), no. 2-3, 131.
- [3] Gy. Elekes, M. Simonovits and E. Szabó, A combinatorial distinction between unit circles and straight lines: how many coincidences can they have? *Combin. Probab. Comput.* 18 (2009), no. 5, 691–705. (Tom Trotter birthday volume)
- [4] György Elekes and Endre Szabó, How to find groups? (And how to use them in Erdős geometry?), submitted.
- [5] M. Simonovits, Personal reminiscences, Gyuri Elekes. This volume.