ΜΑΘΗΜΑΤΙΚΑ.— **Remarks on the Hilbert's 16th problem,** *by Themistocles M. Rassias**. Ἀνεκοινώθη ὑπὸ τοῦ Ἀκαδημαϊκοῦ κ. Φ. Βασιλείου.

1. INTRODUCTION

The problem considered here is to place a bound on the number of regions into which a real algebraic curve divides the real plane $\mathbf{R}^2$ in terms of the degree m of the polynomial defining the curve. This bound is found to be $1 + \frac{m(m+1)}{2}$ (see Theorem 2). This bound will be achieved using a curve made up of m straight lines in general position (see Remark after Theorem 1), and hence is the best possible. In the proof of Theorem 2, an irreducible curve, defined by an irreducible real polynomial of degree m, is considered. The notion of a circuit is introduced. A *circuit* is obtained by starting out along one arc of a real branch and proceeding «naturally». A circuit is a closed path and an irreducible curve divides naturally into a number of circuits and into connected bunches of circuits. The case of a circuit is first studied. The main tool used is Euler's Theorem which states that $v - e + f = 2$, where f is the number of regions determined by a closed path, e is the number of edges of the path, and v (assumed $\geqslant 1$) is the number of vertices. An initial step is to deform the circuit into a closed path having only simple crossings and for such a closed path $e = 2v$. Similar considerations hold for a connected bunch of circuits. The final step is to give a bound on the number of circuits. In the case of a curve without singularities, Harnack's theorem says that $p + 1$ is such a bound, where p is the genus. However, the number of circuits is a birational invariant and thus is p. Since one can desingularize an algebraic curve using a birational transformation, Harnack's bound also holds even if the curve has singularities (see proposition at the end). Straightforward computations now complete the proof. For reducible curves one treats the cases $m = 1$, $m = 2$ and then makes an induction on m.

---

* ΘΕΜΙΣΤΟΚΛΗ ΡΑΣΣΙΑ, **Παρατηρήσεις ἐπὶ τοῦ 16ου προβλήματος τοῦ Hilbert.**

T h e o r e m 1. *The number of regions into which* m *straight lines* $l_1, l_2, l_3, \ldots, l_m$ *can divide the plane,* $\mathbf{R}^2$*, is at most* $1 + [m(m+1)/2]$.

P r o o f. Consider the case $m = 1$, the obviously l_1 divides the plane into two regions. If $m = 2$ then also obviously l_1, l_2 divide the plane into three or four regions. The remaining case is to examine what happens for $m > 2$. To do that we proceed by induction on m. Take

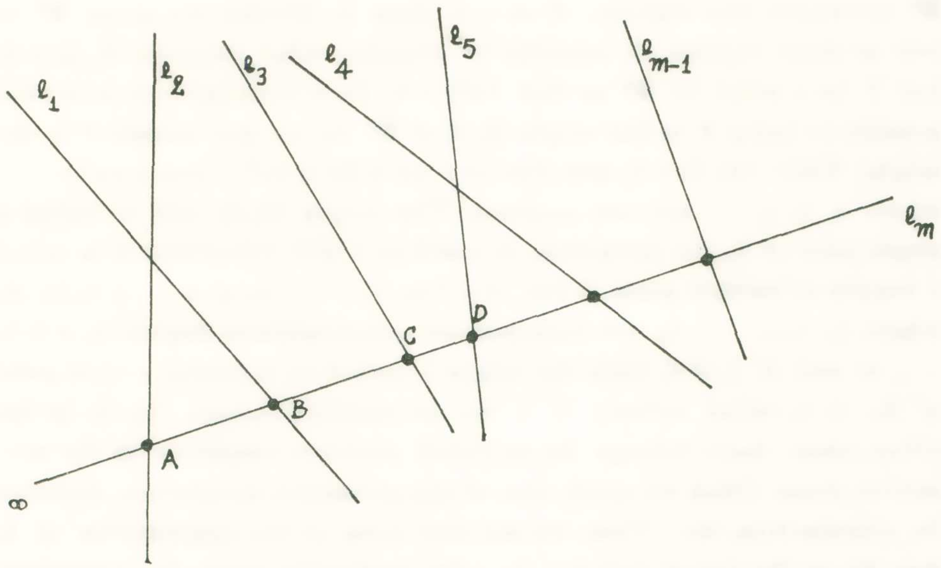


Fig. 1.

$(m-1)$ lines $l_1, l_2, \ldots, l_{m-1}$ among the given m lines. Those $(m-1)$ lines divide the plane into at most $1 + [(m-1)m/2]$ regions, by the induction assumption. Now put down the line l_m. It is seen that l_m cuts $\{l_1, l_2, l_3, \ldots l_{m-1}\}$ in at most $(m-1)$ points A, B, C, ...; and l_m is cut by $\{l_1, l_2, l_3, \ldots l_{m-1}\}$ into at most m pieces, ${}_{\infty}$A, AB, BC, ... (see Figure 1).

Consider the piece CD, say. Removing it diminishes the number of regions by at most one, putting it back increases the number of regions by at most one. Now take away l_m and then put it back. Then we get at most $1 + [m(m-1)/2] + m$ regions, i.e. at most $1 + [m(m+1)/2]$ regions. Q. E. D.

Remark. One can always achieve the number $1 + [m(m+1)/2]$.

Theorem 2. *Let* $f(x, y)$ *be a polynomial of degree* m. *Then the set* $Z_f = \{(x, y) \in \mathbf{R}^2 : f(x, y) = 0\}$ *divides the plane* $\mathbf{R}^2$ *into at most* $1 + [m(m+1)/2]$ *regions.*

Proof. We distinguish two cases, one when the polynomial f is irreducible and two, when f is reducible.

Case 1. Assume f is *irreducible.* If $m = 1$ then Z_f divides the plane $\mathbf{R}^2$ in exactly two regions; if $m = 2$, then Z_f divides the plane $\mathbf{R}^2$ in two or three regions. It remains to examine what happens if $m > 2$. Let P be a point in $\mathbf{R}^2$ so that $f(P) = 0$. By a translation it is always possible to bring P to the origin $(0, 0)$ of $\mathbf{R}^2$. So we may assume P is the origin. Then $f(0, 0) = 0$ and $f(x, y) = ax + by + cx^2 + dxy + ey^2 + \dots$, where a, b, c, ... are real numbers. The origin $(0, 0)$ will be called a *simple point* of Z_f, by definition, if $a \neq 0$ or $b \neq 0$. Otherwise it is called a *singular* or *multiple point.* If $f(x, y) = f_r(x, y) + f_{r+1}(x, y) + \dots + f_m(x, y)$, where $f_r, f_{r+1}, \dots, f_m$ are homogeneous polynomials of degree $r, r+1, \dots, m$ and if $f_r \neq 0$, then the origin is called an (exactly) r-*fold point of* Z_f. It is called *ordinary* if f_r has no multiple factors. To Z_f in the affine plane there belongs its so-called *projective completion* in the projective plane. Often we speak also of this projective completion, defining its singularities, etc. Thus we say that some of the singularities of Z_f may lie on the line at infinity, l_∞. An irreducible curve has something called the *genus,* written p or g. The genus is an integer $\geqslant 0$. If all the singularities are ordinary (in the projective plane), then the genus is given by

$$p = \frac{(m-1)(m-2)}{2} - \frac{\sum_i r_i(r_i - 1)}{2}.$$

Here the curve has an ordinary r_1-fold point, an ordinary r_2-fold point, etc. Even if the singularities are not ordinary, we can talk about singularities *infinitely near others* in such a way that the above formula continues to hold. If these infinitely near singularities are neglected, then we get

$$p \leqslant \frac{(m-1)(m-2)}{2} - \frac{\sum_i r_i(r_i - 1)}{2}.$$

Let us go back to the case that the singularities are ordinary. Some of these singularities may be on l_∞, so neglecting these one gets again

$$p \leqslant \frac{(m-1)\,(m-2)}{2} - \frac{\sum_i r_i\,(r_i-1)}{2}.$$

Thus in any case we get

$$p \leqslant \frac{(m-1)\,(m-2)}{2} - \frac{\sum_i r_i\,(r_i-1)}{2}$$

whether in the affine plane or in the projective plane (and whether singularities are ordinary or not).

Now we are going to explain what is a *circuit*. A real algebraic curve may have isolated points. At a non-isolated point P there will be

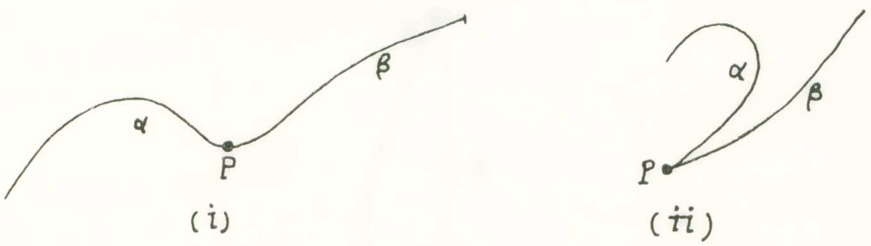


Fig. 2.

centered at least one real branch. A real branch is made up of two arcs terminating at P, say α, β (see Figure 2, (i) and (ii)) (for the theory of algebraic curves, see [4])

Let us enter P say along α, and then emerge along β. Continuing along β we come to a singular point, perhaps P itself. Going into a singular point along one arc of a branch, we agree to leave it along the other arc of the same branch. (This agreement defines the «natural» way to traverse the circuit). Since there are only a finite number of singularities and only a finite number of branches centered at the singularities, we eventually return to the point P along α. The closed path thus traversed is called a *circuit*. For example a circuit may have the appearance as in Figure 3.

Now a projective plane curve is made up of a number of circuits. The circuit may cross itself, as in Figure 3. Out of each point of the circuit there will emerge (locally) an even number of arcs. At an r-fold point (ordinary or not) there emerges at most 2r arcs. Consider a circuit and let P be a point on the circuit. There will be s real branches, $s \leqslant r$, centered at P, and 2s arcs. Each branch is made up of two arcs meeting at P. Number the arcs clockwise as $1, 2, 3, \ldots, 2s$. Now traverse the circuit beginning with arc 1. There is a natural way of traversing the circuit, namely whenever we go into a point along an arc of a branch γ

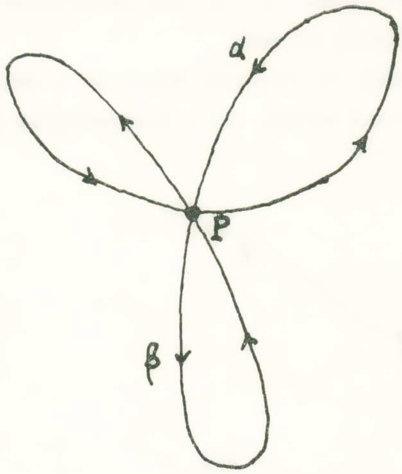
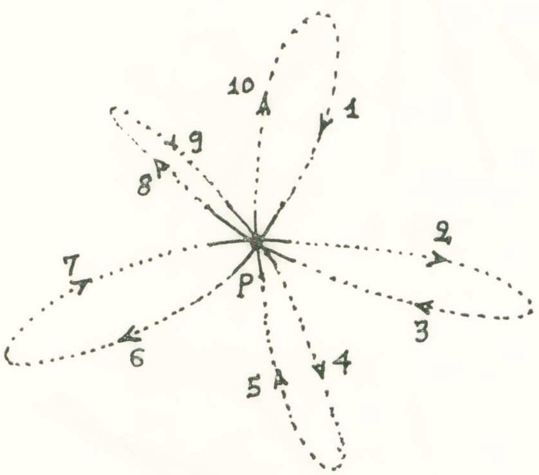


Fig. 3.

we emerge also from the point along the other arc of γ. Assume we have traversed the circuit in a natural way and come into the point P along arc 1. Let the arc 1 be paired with arc i_2. Coming into P along arc 1 we emerge along arc i_2. Now we continue and eventually we come to arc $(1+s)$. We see that either we enter P along arc $(1+s)$ or we emerge from P along arc $(1+s)$. Consider the first case, then we enter P along arc $(1+s)$. Now instead of traversing $i_2, i_3, \ldots, (1+s)$ we traverse this part in the opposite direction, that is we traverse successively $(1+s), \ldots, i_3, i_2$. The result of this is that we now enter along arc 1 and emerge along arc $(1+s)$. In the second case we emerge from the point P along the arc $(1+s)$. Let the arc $(1+s)$ be paired with arc i_h, i.e., $(1+s)$ and i_h are the two arcs of a single branch; then in the natural traversing of the circuit we traverse $1, i_2, i_3, \ldots i_h, 1+s, \ldots$ There is a closed path made up of arcs $i_2, i_3, \ldots, i_h$. We then break up the circuit into two closed paths, namely $i_2, i_3, \ldots, i_h$ and $1, 1+s, \ldots$ We now have one or two closed paths and a way of traversing each of them. Let i be an arc at P different from 1 and $(1+s)$. In the traversing we now have that we enter along arc i and emerge along some arc δ and we continue. If the arc δ is the arc $(1+s)$ mod 2s then we are done, otherwise one argues as before, namely one continues with $\delta, \ldots$



1, 6, 7, 2, 3, 8, 9, 4, 5, 10

Fig. 4.

eventually coming to arc $(i+s)$ mod 2s. We distinguish two cases: in the first case the arc $(i+s)$ mod 2s enters P and in the second case it emerges from P, etc. Eventually we have broken up the original circuit into several closed paths and we have replaced the natural traversing of the circuit by traversings of the closed paths and in such a way that when we enter the point P along the arc i we emerge along the arc $(i+s)$ mod 2s. We carry out this procedure for all circuits and for all singular points. The situation here looks as in Figure 4, where singularities may occur in the dotted parts.

So locally at P it is very much like a number of straigh lines (see Figure 5).

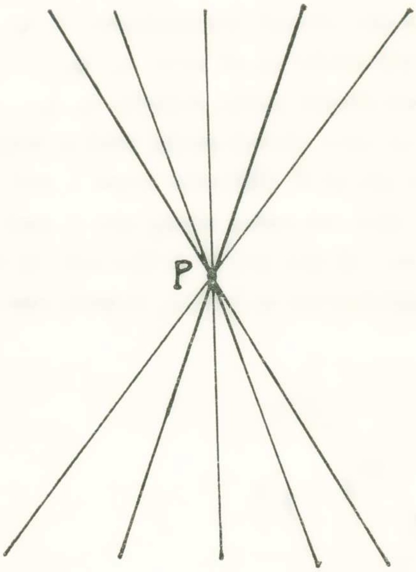


Fig. 5.

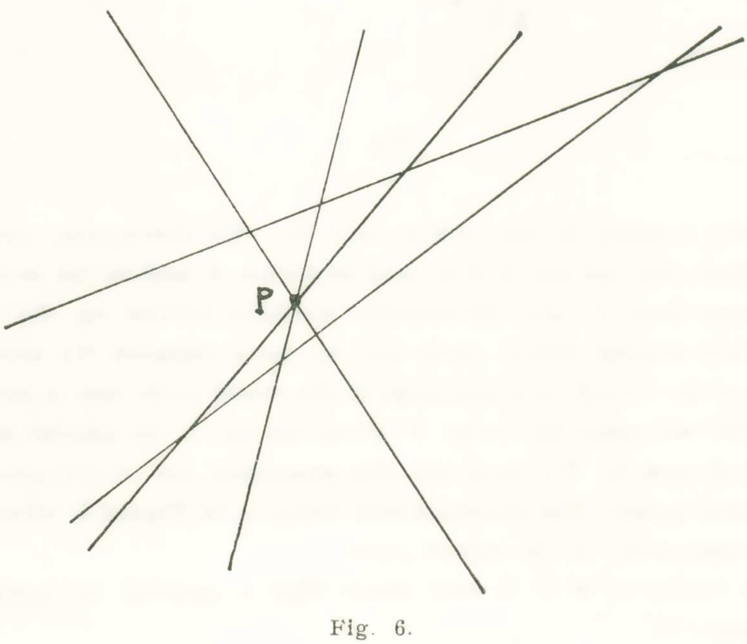


Fig. 6.

Thus if we separate these lines a little as in Figure 6, the number of regions does not diminish, rather it grows. To see this take a little circle about the point P, so small that the arcs emerging from P and belonging to our circuit do not meet again in the circle. Let the arcs meet the circle in points $A_1, A_2, A_3, \ldots, A_{2s}$ (see Figure 7).

We have paths connecting A_i and $A_{(i+s) \bmod 2s}$ namely A_i to P and then P to $A_{(i+s) \bmod 2s}$. Replace the pair from A_i to P and then to $A_{(i+s) \bmod 2s}$ by another simple path within the little circle for each i and in such a way that no three of the new paths meet within the circle. (See Figure 8).

The above considerations for a circuit clearly hold also for any closed path. Thus a closed path, like a circuit, is the union of a finite

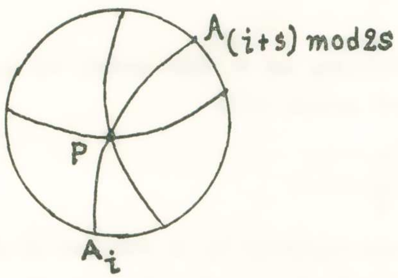


Fig. 7.

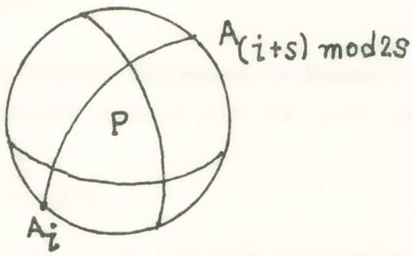


Fig. 8.

number of arcs $A_1 A_2, A_2 A_3, \ldots, A_n A_1$, no two of which meet except at the points A_i; from each A_i there emerge an even number of arcs (which we number clockwise as $1, 2, 3, \ldots, 2s_i$); and if one is given a way of traversing this path, say, $A_1 A_2, A_2 A_3, \ldots, A_n A_1$, then one has a «natural» pairing of the arcs emerging from any A_i (namely, one pairs $A_{i-1} A_i$ and $A_i A_{i+1}$); and one can find a traversing of the path such that at each A_i, when one enters along arc j one emerges along arc $(j+s_i) \bmod 2s_i$. So we may assume that locally at each A_i, the path has the appearance of a number of straight lines; and then we can separate the lines as previously explained. The same result holds for any finite connected union of closed paths any two of which meet in only a finite number of points, since such a union itself is a closed path. And it therefore holds also for a connected bunch of circuits of our algebraic curve.

If P is an r-fold point (ordinary or not) then there are s real branches centered at P and this r-fold point has been replaced by $[s(s-1)/2]$ simple crossings. So in a computation involving an r-fold point (ordinary or not) we can think of it as $[r(r-1)/2]$ (or fewer) simple crossings. By Euler's theorem for a connected graph in the plane $\mathbf{R}^2$, with at least one vertex: $v - e + f = 2$, where v is the number of vertices, e is the number of edges and f is the number of faces or regions. For example in Figure 9 $v = 1$, $e = 2$, $f = 3$ and $v - e + f = 1 - 2 + 3 = 2$.

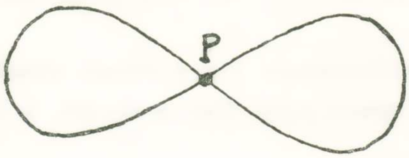


Fig. 9.

For a bunch of connected circuits involving an r_1-fold point, an r_2-fold point, etc., we will have a connected graph with

$$v \leqslant \frac{\sum_i r_i(r_i - 1)}{2}.$$

Assuming now that our curve has been replaced by a number of closed paths having only simple crossings, one finds $e = 2v$. In fact, if $v = 1$, then the path has the form of Figure 9 and $e = 2$, so here $e = 2v$. Now we make an induction on v, assuming $v > 1$. From each vertex there will emerge 3 or 4 arcs (or «edges»); if 3, we will say that there is a loop. If there are no loops, one finds $2e = 4v$, since there are 4 edges for each vertex, but sach edge is counted twice; i. e., $e = 2v$. If there is a loop, say at P, remove the loop, and the vertex P, and merge the other two edges emerging from P. So one has diminished e by 2 and v by 1. So $e - 2 = 2(v - 1)$, by induction, so again $e = 2v$. Thus from $v - e + f = 2$ we get $f = 2 + e - v = 2 + 2v - v = 2 + v$ and so $f = 2 + v$. This holds also if $v = 0$. If there are π connected bunches of circuits, we get

$$f = 1 + \sum_{i=1}^{\pi}(1 + k_i) \leqslant 1 + \pi + \sum_{i,j} \frac{r_{ij}(r_{ij} - 1)}{2},$$

where π is the number of connected bunches of circuits and k_i is the number of vertices on a bunch of connected circuits. Here we first counted only $(1+k_i)$ instead of $(2+k_i)$ regions, since all the circuits share a region, and then we added 1. It is also true that

$$k_i \leqslant \frac{\sum_j r_{ij}(r_{ij}-1)}{2}$$

for the multiple points (ordinary or not) on a connected bunch of circuits i. Now for a non-singular curve, not necessarily planar, there are at most $(p+1)$ circuits (and a fortiori at most $(p+1)$ connected bunches of circuits). This is in projective space. This formula also holds for curves with singularities (see proposition at the end). Also because we are in the affine plane, we add m regions because of the way the curve cuts l_∞. So altogether we get at most

$$\left\{1+\left[\frac{(m-1)(m-2)}{2}-\frac{\sum_{i,j} r_{ij}(r_{ij}-1)}{2}+1\right]+\frac{\sum_{i,j} r_{ij}(r_{ij}-1)}{2}+m\right\}$$

regions or $\left[\frac{(m-1)(m-2)}{2}+m+2\right]$ regions. But

$$\frac{(m-1)(m-2)}{2}+m+2 \leqslant 1+\frac{m(m+1)}{2}$$

for $m \geqslant 2$ which happens, since we have assumed $m>2$. We have examined the cases $m=1$ and $m=2$ separately. This finishes the proof for f an irreducible polynomial of degree m.

Case 2. Assume now that f is a *reducible* polynomial of degree m. We may suppose that f has no multiple factors since in replacing a factor $f_i^{c_i}$, $c_i>1$, by f_i, the degree of f is lowered. Now, if f has a linear factor l then $f=l\cdot H$ where H can be either irreducible or reducible. Then we argue as in Theorem 1 to get the result by induction on m. Also if f has a quadratic factor, we argue in nearly the same way to get the result. So suppose $f=f_1 f_2 f_3 \ldots f_s$, with $f_i=f_i(x, y)$ irreducible, non-associates (i.e. $f_i \neq c f_j$ for any constant c) and $\deg f_i>2$, $\deg f_i=m_i$. Here $s \leqslant \frac{m}{3}$. Now we consider the case of bunches of connected circuits of the polynomial f_i and argue as before. The new thing is that there will be $\sum_{i \neq j} m_i m_j$ new multiple points to think about. One argues about these very much like about the multiple points of an irreducible polynomial f_i. Altogether there will be multiplicities summing to $\sum_{i \neq j} m_i m_j + \sum_{i,j} \frac{r_{ij}(r_{ij}-1)}{2}$ instead of to $\sum_{i,j} \frac{r_{ij}(r_{ij}-1)}{2}$. Here the r_{ij} are the multiplicities of the multiple points on Z_{f_i}.

[For example, suppose a 3-fold point of Z_{f_1} and a 4-fold point of Z_{f_2} occur at the same point P, and say $Z_{f_3}, \ldots, Z_{f_s}$ do not pass through

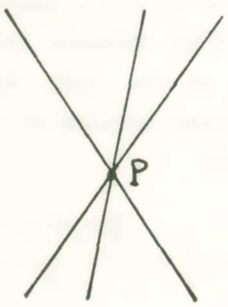


Fig. 10.

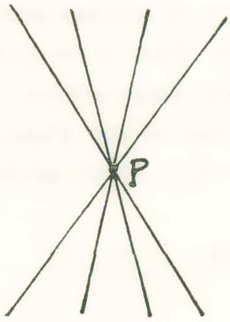


Fig. 11.

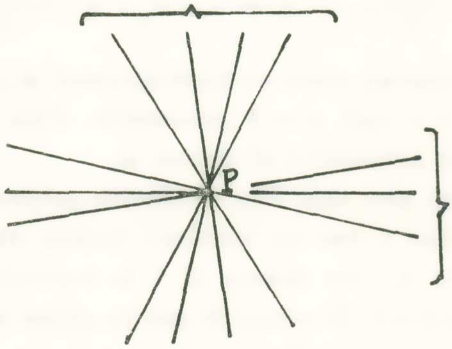


Fig. 12.

P. Then topologically, from the real point of view, we look at Z_{f_1} at P as in Figure 10.

There may be *less* than three real pieces as shown. We look at Z_{f_2} at the point P as in Figure 11.

There may be *less* than four real pieces as shown Altogether we have as in Figure 12.

If Z_{f_1} has an r'-fold point at P and Z_{f_2} has an r''-fold point at P, then Z_{f_1} and Z_{f_2} meet with multiplicity $\geqslant r'r''$ at P (by a known theorem). Now separate the lines a little as in Figure 13.

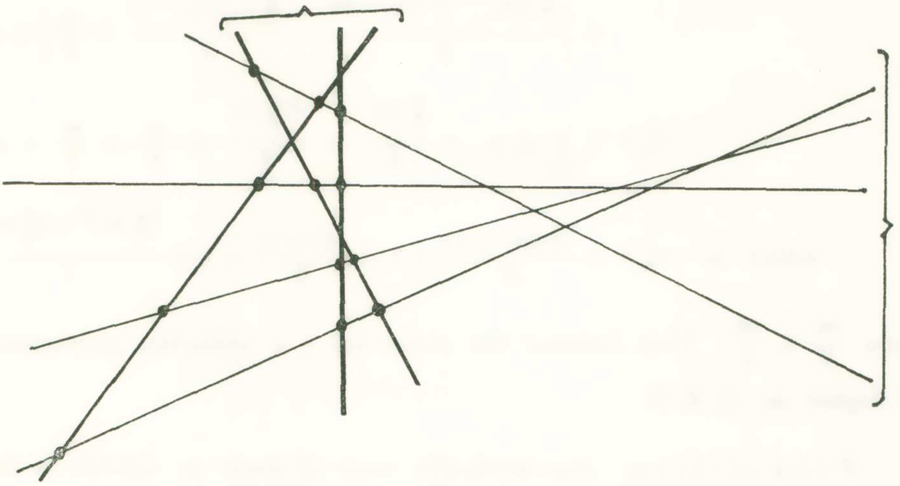


Fig. 13.

We get $[3(3-1)/2]$ points for Z_{f_1}, $[4(4-1)/2]$ points for Z_{f_2} and $3 \cdot 4$ points for the intersection of both Z_{f_1} and Z_{f_2}. If several of the Z_{f_i}, perhaps more than 2, meet at P, then the equivalent number of double points is at least equal to the sum of the number for each Z_{f_i} plus the number for each pair Z_{f_i}, Z_{f_j} $(i \neq j)$, i.e. if Z_{f_i} has at P a d_i-fold point, then the equivalent number of double points is $\geqslant \sum_i \frac{d_i(d_i-1)}{2} + \sum_{i \neq j} d_i d_j$.]

This time we have $\sum_1 (p_i + 1)$ connected bunches of circuits at most; or $(\sum_1 p_i + s)$ connected bunches of circuits at most; or $\left(\sum_i p_i + \frac{m}{3}\right)$ connected bunches of circuits at most. We get $\sum_i \frac{r_i(r_i-1)}{2} + \sum_{i \neq j} m_i m_j$ multiple points at most (think of them as ordinary double points). Then the number of regions is at most

$$1 + \sum_i (1 + k_i) + m \leqslant 1 + \sum_i k_i + \left[\sum_i p_i + \frac{m}{3}\right] + m$$

$$\leqslant 1 + \sum_{i,j} \frac{r_{ij}(r_{ij} - 1)}{2} + \sum_{i \neq j} m_i m_j$$

$$+ \left[\frac{\sum_i (m_i - 1)(m_i - 2)}{2} - \frac{\sum_{i,j} r_{ij}(r_{ij} - 1)}{2} + \frac{m}{3}\right] + m$$

$$\leqslant 1 + \sum_{i \neq j} m_i m_j + \frac{\sum_i m_i^2}{2} - \frac{3\sum_i m_i}{2} + \frac{m}{3} + \frac{m}{3} + m,$$

which is $\leqslant 1 + \frac{m(m+1)}{2} = 1 + \frac{m^2 + m}{2} = 1 + \frac{(\sum_i m_i)^2 + (\sum_i m_i)}{2}$

since $\frac{m}{6} \leqslant \frac{m}{2}$. This finishes the proof for f a reducible polynomial of degree m. Q. E. D.

Proposition. *Any irreducible curve of genus* p, *whether it has singularities or not, has at most* $(p + 1)$ *circuits.*

Proof. If the algebraic curve is free of singularities this is known [2; pp. 257 - 258] as Harnack's theorem. But even in the case of singularities, $(p + 1)$ will be the bound as we will show. In the case there are singularities, one can apply a birational transformation to our curve to get a transform without singularities. This is known and it is also known that there is a 1—1 correspondence between the branches of the two curves, and hence there is also a 1 — 1 correspondence between the circuits of the two curves. So the number of circuits is a birational invariant; and since the genus is also a birational invariant, weget that the number of circuits is at most $p + 1$ in general. Q. E. D.

Some other very interesting results concerning the localization of zeros of polynomials have been investigated by S. P. Zervos (see, for example, [6]).

Remark. The above result (Theorem 2) will be applied in a subsequent paper to find the maximum number of regions into which a spherical harmonic divides the sphere S^2. This sharpens the Courant's nodal line theorem [1] for spherical harmonics which is useful in the study of the Morse-Smale Index [3], [5] of the Jacobi operator corresponding to a variational problem.

Acknowledgments. It is my pleasure to express my gratitude to Professors Abraham Seidenberg and Stephen Smale for helpful conversations.

ΠΕΡΙΛΗΨΙΣ

Ἡ παροῦσα ἐργασία ἀναφέρεται εἰς τὴν εὕρεσιν τοῦ μεγίστου ἀριθμοῦ χωρίων, εἰς τὸν ὁποῖον δοθεῖσα πραγματικὴ ἀλγεβρικὴ καμπύλη διαιρεῖ τὸ πραγματικὸν ἐπίπεδον $\mathbf{R}^2$ συναρτήσει τοῦ βαθμοῦ m τοῦ πολυωνύμου, τὸ ὁποῖον ὁρίζει τὴν καμπύλην. Αὐτὸς ὁ μέγιστος ἀριθμὸς χωρίων εὑρέθη ὅτι ἰσοῦται μὲ $1 + [m(cm+1)/2]$.

Τὸ ἀνωτέρω συμπέρασμα παρέχει λύσιν μιᾶς ἐνδιαφερούσης περιπτώσεως τοῦ περιφήμου 16ου προβλήματος τοῦ Hilbert, τὸ ὁποῖον οὗτος ἔθεσεν εἰς τὸ παγκόσμιον συνέδριον Μαθηματικῶν εἰς Παρισίους τὸ ἔτος 1900. Ἐφαρμογαὶ τοῦ ἀνωτέρω ἀποτελέσματος εἰς τὸν ὑπολογισμὸν τοῦ Morse - Smale index, διὰ τὸν τελεστὴν τοῦ Jacobi, ὁ ὁποῖος ἀντιστοιχεῖ εἰς ἕνα πρόβλημα τοῦ Λογισμοῦ τῶν Μεταβολῶν (Calculus of Variations), δύνανται νὰ δοθοῦν.

REFERENCES

1. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. 1, Interscience, New York (1953).
2. S. Lang, Introduction to Algebraic Geometry, Interscience Publishers, Inc., New York (1958).
3. T. M. Rassias, Sur la multiplicité du premier bord conjugué d'une hypersurface minimale de $\mathbf{R}^n$, $n \geqslant 3$ Comptes Rendus, Acad. Sci. Paris t. 284, Série A, 497 - 499 (1977).
4. A. Seidenberg, Elements of the Theory of Algebraic Curves, Addison-Wesley Publishing Co., Reading, Mass (1968).
5. S. Smale, On the Morse Index Theorem, J. Math. Mech. 14, 1049 - 1056 (1965).
6. S. P. Zervos, Aspects modernes de la localisation des zéros des polynomes d'une variable, Ann. Sci. Éc. Norm. Sup., 3e Séries, t. 77, 303 à 410 (1960).