INTRODUCTION TO TOPOLOGY OF REAL ALGEBRAIC VARIETIES

OLEG VIRO

1. The Early Topological Study of Real Algebraic Plane Curves

1.1. **Basic Definitions and Problems.** A curve (at least, an algebraic curve) is something more than just the set of points which belong to it. There are many ways to introduce algebraic curves. In the elementary situation of real plane projective curves the simplest and most convenient is the following definition, which at first glance seems to be overly algebraic.

By a *real projective algebraic plane curve*[1] *of degree* m we mean a homogeneous real polynomial of degree m in three variables, considered up to constant factors. If a is such a polynomial, then the equation $a(x_0, x_1, x_2) = 0$ defines the *set of real points of the curve* in the real projective plane $\mathbb{R}P^2$. We let $\mathbb{R}A$ denote the set of real points of the curve A. Following tradition, we shall also call this set a curve, avoiding this terminology only in cases where confusion could result.

A point $(x_0 : x_1 : x_2) \in \mathbb{R}P^2$ is called a (real) *singular point* of the curve A if $(x_0, x_1, x_2) \in \mathbb{R}^3$ is a critical point of the polynomial a which defines the curve. The curve A is said to be (*real*) *nonsingular* if it has no real singular points. The set of real points of a nonsingular real projective plane curve is a smooth closed one-dimensional submanifold of the projective plane.

In the topology of nonsingular real projective algebraic plane curves, as in other similar areas, the first natural questions that arise are classification problems.

1.1.A (Topological Classification Problem)**.** *Up to homeomorphism, what are the possible sets of real points of a nonsingular real projective algebraic plane curve of degree m?*

[1]Of course, the full designation is used only in formal situations. One normally adopts an abbreviated terminology. We shall say simply a *curve* in contexts where this will not lead to confusion.

1.1.B (Isotopy Classification Problem)**.** *Up to homeomorphism, what are the possible pairs* $(\mathbb{R}P^2, \mathbb{R}A)$ *where* A *is a nonsingular real projective algebraic plane curve of degree* m*?*

It is well known that the components of a closed one-dimensional manifold are homeomorphic to a circle, and the topological type of the manifold is determined by the number of components; thus, the first problem reduces to asking about the number of components of a curve of degree m. The answer to this question, which was found by Harnack [Har-76] in 1876, is described in Sections 1.6 and 1.8 below.

The second problem has a more naive formulation as the question of how a nonsingular curve of degree m can be situated in $\mathbb{R}P^2$. Here we are really talking about the isotopy classification, since any homeomorphism $\mathbb{R}P^2 \to \mathbb{R}P^2$ is isotopic to the identity map. At present the second problem has been solved only for $m \leq 7$. The solution is completely elementary when $m \leq 5$: it was known in the last century, and we shall give the result in this section. But before proceeding to an exposition of these earliest achievements in the study of the topology of real algebraic curves, we shall recall the isotopy classification of closed one-dimensional submanifolds of the projective plane.

1.2. **Digression: the Topology of Closed One-Dimensional Submanifolds of the Projective Plane.** For brevity, we shall refer to closed one-dimensional submanifolds of the projective plane as *topological plane curves*, or simply *curves* when there is no danger of confusion.

A connected curve can be situated in $\mathbb{R}P^2$ in two topologically distinct ways: *two-sidedly*, i.e., as the boundary of a disc in $\mathbb{R}P^2$, and *one-sidedly*, i.e., as a projective line. A two-sided connected curve is called an *oval*. The complement of an oval in $\mathbb{R}P^2$ has two components, one of which is homeomorphic to a disc and the other homeomorphic to a Möbius strip. The first is called the *inside* and the second is called the *outside*. The complement of a connected one-sided curve is homeomorphic to a disc.

Any two one-sided connected curves intersect, since each of them realizes the nonzero element of the group $H_1(\mathbb{R}P^2; \mathbb{Z}_2)$, which has nonzero self-intersection. Hence, a topological plane curve has at most one one-sided component. The existence of such a component can be expressed in terms of homology: it exists if and only if the curve represents a nonzero element of $H_1(\mathbb{R}P^2; \mathbb{Z}_2)$. If it exists, then we say that the whole curve is *one-sided*; otherwise, we say that the curve is *two-sided*.

Two disjoint ovals can be situated in two topologically distinct ways: each may lie outside the other one—i.e., each is in the outside component of the complement of the other—or else they may form an *injective*

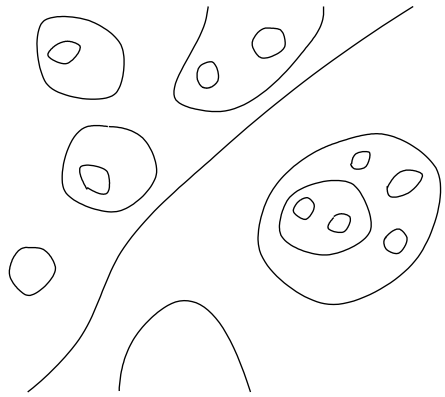

FIGURE 1

pair, i.e., one of them is in the inside component of the complement of the other—in that case, we say that the first is the *inner* oval of the pair and the second is the *outer* oval. In the latter case we also say that the outer oval of the pair *envelopes* the inner oval.

A set of h ovals of a curve any two of which form an injective pair is called a *nest of depth h*.

The pair $(\mathbb{R}P^2, X)$, where X is a topological plane curve, is determined up to homeomorphism by whether or not X has a one-sided component and by the relative location of each pair of ovals. We shall adopt the following notation to describe this. A curve consisting of a single oval will be denoted by the symbol $\langle 1\rangle$. The empty curve will be denoted by $\langle 0\rangle$. A one-sided connected curve will be denoted by $\langle J\rangle$. If $\langle A\rangle$ is the symbol for a certain two-sided curve, then the curve obtained by adding a new oval which envelopes all of the other ovals will be denoted by $\langle 1\langle A\rangle\rangle$. A curve which is a union of two disjoint curves $\langle A\rangle$ and $\langle B\rangle$ having the property that none of the ovals in one curve is contained in an oval of the other is denoted by $\langle A \amalg B\rangle$. In addition, we use the following abbreviations: if $\langle A\rangle$ denotes a certain curve, and if a part of another curve has the form $A \amalg A \amalg \cdots \amalg A$, where A occurs n times, then we let $n \times A$ denote $A \amalg \cdots \amalg A$. We further write $n \times 1$ simply as n.

When depicting a topological plane curve one usually represents the projective plane either as a disc with opposite points of the boundary identified, or else as the compactification of $\mathbb{R}^2$, i.e., one visualizes the curve as its preimage under either the projection $D^2 \to \mathbb{R}P^2$ or the inclusion $\mathbb{R}^2 \to \mathbb{R}P^2$. In this book we shall use the second method. For example, 1.2 shows a curve corresponding to the symbol $\langle J \amalg 1 \amalg 2\langle 1\rangle \amalg 1\langle 2\rangle \amalg 1\langle 3 \amalg 1\langle 2\rangle\rangle\rangle$.

1.3. **Bézout's Prohibitions and the Harnack Inequality.** The most elementary prohibitions, it seems, are the topological consequences of Bézout's theorem. In any case, these were the first prohibitions to be discovered.

1.3.A (Bézout's Theorem (see, for example, [Wal-50], [Sha-77]))**.** *Let A_1 and A_2 be nonsingular curves of degree m_1 and m_2. If the set $\mathbb{R}A_1 \cap \mathbb{R}A_2$ is finite, then this set contains at most $m_1 m_2$ points. If, in addition, $\mathbb{R}A_1$ and $\mathbb{R}A_2$ are transversal to one another, then the number of points in the intersection $\mathbb{R}A_1 \cap \mathbb{R}A_2$ is congruent to $m_1 m_2$ modulo* 2*.*

1.3.B. **Corollary** (1)**.** *A nonsingular plane curve of degree m is one-sided if and only if m is odd. In particular, a curve of odd degree is nonempty.*

In fact, in order for a nonsingular plane curve to be two-sided, i.e., to be homologous to zero $\mod 2$, it is necessary and sufficient that its intersection number with the projective line be zero $\mod 2$. By Bézout's theorem, this is equivalent to the degree being even. □

1.3.C. **Corollary** (2)**.** *The number of ovals in the union of two nests of a nonsingular plane curve of degree m does not exceed $m/2$. In particular, a nest of a curve of degree m has depth at most $m/2$, and if a curve of degree m has a nest of depth $[m/2]$, then it does not have any ovals not in the nest.*

To prove Corollary 2 it suffices to apply Bézout's theorem to the curve and to a line which passes through the insides of the smallest ovals in the nests. □

1.3.D. **Corollary** (3)**.** *There can be no more than m ovals in a set of ovals which is contained in a union of ≤ 5 nests of a nonsingular plane curve of degree m and which does not contain an oval enveloping all of the other ovals of the set.*

To prove Corollary 3 it suffices to apply Bézout's theorem to the curve and to a conic which passes through the insides of the smallest ovals in the nests. □

One can give corollaries whose proofs use curves of higher degree than lines and conics (see Section 3.8). The most important of such results is Harnack's inequality.

1.3.E. **Corollary** (4 (Harnack Inequality [Har-76]))**.** *The number of components of a nonsingular plane curve of degree m is at most $\frac{(m-1)(m-2)}{2}+1$.*

The derivation of Harnack Inequality from Bézout's theorem can be found in [Har-76], and also [Gud-74]. However, it is possible to prove Harnack Inequality without using Bézout's theorem; see, for example, [Gud-74], [Wil-78] and Section 3.2 below.

1.4. **Curves of Degree ≤ 5.** If $m \leq 5$, then it is easy to see that the prohibitions in the previous subsection are satisfied only by the following isotopy types.

TABLE 1

m	Isotopy types of nonsingular plane curves of degree m
1	$\langle J \rangle$
2	$\langle 0 \rangle$, $\langle 1 \rangle$
3	$\langle J \rangle$, $\langle J \amalg 1 \rangle$
4	$\langle 0 \rangle$, $\langle 1 \rangle$, $\langle 2 \rangle$, $\langle 1\langle 1 \rangle\rangle$, $\langle 3 \rangle$, $\langle 4 \rangle$
5	$\langle J \rangle$, $\langle J \amalg 1 \rangle$, $\langle J \amalg 2 \rangle$, $\langle J \amalg 1\langle 1 \rangle\rangle$, $\langle J \amalg 3 \rangle$, $\langle J \amalg 4 \rangle$, $\langle J \amalg 5 \rangle$, $\langle J \amalg 6 \rangle$

For $m \leq 3$ the absence of other types follows from *1.3.B* and *1.3.C*; for $m = 4$ it follows from *1.3.B*, *1.3.C* and *1.3.D*, or else from *1.3.B*, *1.3.C* and *1.3.E*; and for $m = 5$ it follows from *1.3.B*, *1.3.C* and *1.3.E*. It turns out that it is possible to realize all of the types in Table 1; hence, we have the following theorem.

1.4.A. ISOTOPY CLASSIFICATION OF NONSINGULAR REAL PLANE PROJECTIVE CURVES OF DEGREE ≤ 5. *An isotopy class of topological plane curves contains a nonsingular curve of degree $m \leq 5$ if and only if it occurs in the m-th row of Table 1.*

The curves of degree ≤ 2 are known to everyone. Both of the isotopy types of nonsingular curves of degree 3 can be realized by small perturbations of the union of a line and a conic which intersect in two real points (Figure 2). One can construct these perturbations by replacing the left side of the equation $cl = 0$ defining the union of the conic C and the line L by the polynomial $cl + \varepsilon l_1 l_2 l_3$, where $l_i = 0$, $i = 1, 2, 3$, are the equations of the lines shown in 2, and ε is a nonzero real number which is sufficiently small in absolute value.

It will be left to the reader to prove that one in fact obtains the curves in Figure 2 as a result; alternatively, the reader can deduce this fact from the theorem in the next subsection.

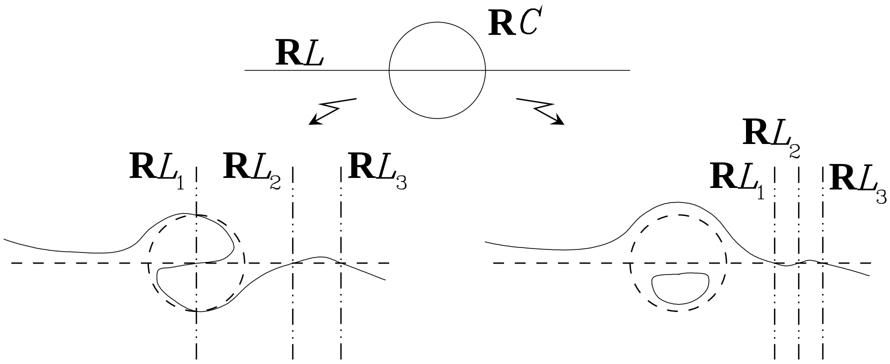

FIGURE 2

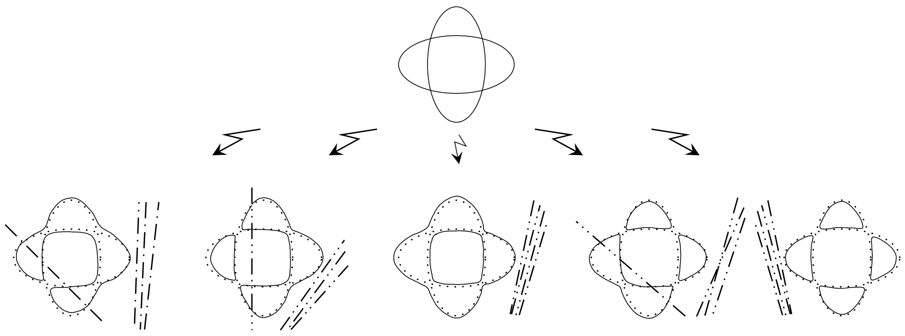

FIGURE 3

The isotopy types of nonempty nonsingular curves of degree 4 can be realized in a similar way by small perturbations of a union of two conics which intersect in four real points (Figure 3). An empty curve of degree 4 can be defined, for example, by the equation $x_0^4+x_1^4+x_2^4=0$.

All of the isotopy types of nonsingular curves of degree 5 can be realized by small perturbations of the union of two conics and a line, shown in Figure 4. $\square$

For the isotopy classification of nonsingular curves of degree 6 it is no longer sufficient to use this type of construction, or even the prohibitions in the previous subsection. See Section 1.13 and **??**.

1.5. **The Classical Method of Constructing Nonsingular Plane Curves.** All of the classical constructions of the topology of nonsingular plane curves are based on a single construction, which I will call

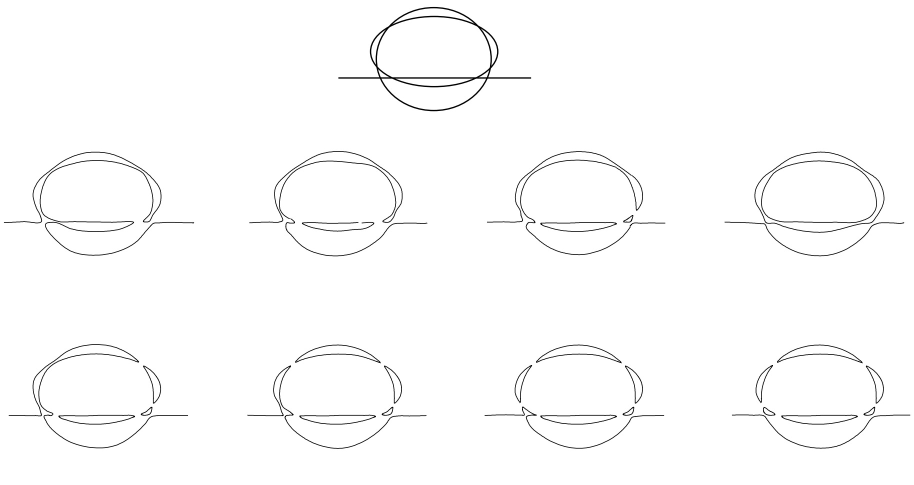

FIGURE 4

classical small perturbation. Some special cases were given in the previous subsection. Here I will give a detailed description of the conditions under which it can be applied and the results.

We say that a real singular point $\xi = (\xi_0 : \xi_1 : \xi_2)$ of the curve A is an *intersection point of two real transversal branches*, or, more briefly, a *crossing*,[2] if the polynomial a defining the curve has matrix of second partial derivatives at the point (ξ_0, ξ_1, ξ_2) with both a positive and a negative eigenvalue, or, equivalently, if the point ξ is a nondegenerate critical point of index 1 of the functions $\{x \in \mathbb{R}P^2 | x_i \neq 0\} \to \mathbb{R}\ x \mapsto a(x)/x_i \deg a$ for i with $\xi_i \neq 0$. By Morse lemma (see, e.g. [Mil-69]) in a neighborhood of such a point the curve looks like a union of two real lines. Conversely, if $\mathbb{R}A_1, \dots, \mathbb{R}A_k$ are nonsingular mutually transverse curves no three of which pass through the same point, then all of the singular points of the union $\mathbb{R}A_1 \cup \dots \cup \mathbb{R}A_k$ (this is precisely the pairwise intersection points) are crossings.

1.5.A (Classical Small Perturbation Theorem (see Figure 5))**.** *Let A be a plane curve of degree m all of whose singular points are crossings, and let B be a plane curve of degree m which does not pass through the singular points of A. Let U be a regular neighborhood of the curve $\mathbb{R}A$ in $\mathbb{R}P^2$, represented as the union of a neighborhood U_0 of the set of singular points of A and a tubular neighborhood U_1 of the submanifold $\mathbb{R}A \smallsetminus U_0$ in $\mathbb{R}P^2 \smallsetminus U_0$.*

Then there exists a nonsingular plane curve X of degree m such that:

[2] Sometimes other names are used. For example: a node, a point of type A_1 with two real branches, a nonisolated nondegenerate double point.

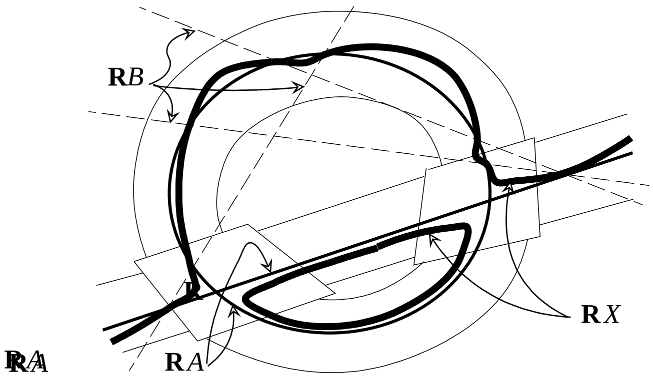

FIGURE 5

(1) $\mathbb{R}X \subset U$.

(2) For each component V of U_0 there exists a homeomorphism $hV \to D^1 \times D^1$ such that $h(\mathbb{R}A \cap V) = D^1 \times 0 \cup 0 \times D^1$ and $h(\mathbb{R}X \cap V) = \{(x, y) \in D^1 \times D^1 | xy = 1/2\}$.

(3) $\mathbb{R}X \smallsetminus U_0$ is a section of the tubular fibration $U_1 \to \mathbb{R}A \smallsetminus U_0$.

(4) $\mathbb{R}X \subset \{(x_0 : x_1 : x_2) \in \mathbb{R}P^2 | a(x_0, x_1, x_2)b(x_0, x_1, x_2) \leq 0\}$, where a and b are polynomials defining the curves A and B.

(5) $\mathbb{R}X \cap \mathbb{R}A = \mathbb{R}X \cap \mathbb{R}B = \mathbb{R}A \cap \mathbb{R}B$.

(6) If $p \in \mathbb{R}A \cap \mathbb{R}B$ is a nonsingular point of B and $\mathbb{R}B$ is transversal to $\mathbb{R}A$ at this point, then $\mathbb{R}X$ is also transversal to $\mathbb{R}A$ at the point.

There exists $\varepsilon > 0$ such that for any $t \in (0, \varepsilon]$ the curve given by the polynomial $a + tb$ satisfies all of the above requirements imposed on X.

It follows from (1)–(3) that for fixed A the isotopy type of the curve $\mathbb{R}X$ depends on which of two possible ways it behaves in a neighborhood of each of the crossings of the curve A, and this is determined by condition (4). Thus, conditions (1)–(4) characterize the isotopy type of the curve $\mathbb{R}X$. Conditions (4)–(6) characterize its position relative to $\mathbb{R}A$.

We say that the curves defined by the polynomials $a + tb$ with $t \in (0, \varepsilon]$ are obtained by *small perturbations of A directed to the curve B*. It should be noted that the curves A and B do not determine the isotopy type of the perturbed curves: since both of the polynomials b and $-b$ determine the curve B, it follows that the polynomials $a - tb$ with small $t > 0$ also give small perturbations of A directed to B. But these curves are not isotopic to the curves given by $a + tb$ (at least not in U), if the curve A actually has singularities.

Proof. Proof of Theorem *1.5.A* We set $x_t = a + tb$. It is clear that for any $t \neq 0$ the curve X_t given by the polynomial x_t satisfies conditions (5) and (6), and if $t > 0$ it satisfies (4). For small $|t|$ we obviously have $\mathbb{R}X_t \subset U$. Furthermore, if $|t|$ is small, the curve $\mathbb{R}X_t$ is nonsingular at the points of intersection $\mathbb{R}X_t \cap \mathbb{R}B = \mathbb{R}A \cap \mathbb{R}B$, since the gradient of x_t differs very little from the gradient of a when $|t|$ is small, and the latter gradient is nonzero on $\mathbb{R}A \cap \mathbb{R}B$ (this is because, by assumption, B does not pass through the singular points of A). Outside $\mathbb{R}B$ the curve $\mathbb{R}X_t$ is a level curve of the function a/b. On $\mathbb{R}A \smallsetminus \mathbb{R}B$ this level curve has critical points only at the singular points of $\mathbb{R}A$, and these critical points are nondegenerate. Hence, for small t the behavior of $\mathbb{R}X_t$ outside $\mathbb{R}B$ is described by the implicit function theorem and Morse Lemma (see, for example, [Mil-69]); in particular, for small $t \neq 0$ this curve is nonsingular and satisfies conditions (2) and (3). Consequently, there exists $\varepsilon > 0$ such that for any $t \in (0, \varepsilon]$ the curve $\mathbb{R}X_t$ is nonsingular and satisfies (1)–(6). $\square$

1.6. **Harnack Curves.** In 1876, Harnack [Har-76] not only proved the inequality *1.3.E* in Section 1.3, but also completed the topological classification of nonsingular plane curves by proving the following theorem.

1.6.A (Harnack Theorem)**.** *For any natural number m and any integer c satisfying the inequalities*

$$\frac{1-(-1)^m}{2} \leq c \leq \frac{m^2 - 3m + 4}{2}, \tag{1}$$

there exists a nonsingular plane curve of degree m consisting of c components.

The inequality on the right in 1 is Harnack Inequality. The inequality on the left is part of Corollary 1 of Bézout's theorem (see Section *1.3.B*). Thus, Harnack Theorem together with theorems *1.3.B* and *1.3.E* actually give a complete characterization of the set of topological types of nonsingular plane curves of degree m, i.e., they solve problem *1.1.A*.

Curves with the maximum number of components (i.e., with $(m^2 - 3m + 4)/2$ components, where m is the degree) are called M-curves. Curves of degree m which have $(m^2 - 3m + 4)/2 - a$ components are called $(M - a)$-curves. We begin the proof of Theorem *1.6.A* by establishing that the Harnack Inequality *1.3.B* is best possible.

*1.6.B***.** *For any natural number m there exists an M-curve of degree m.*

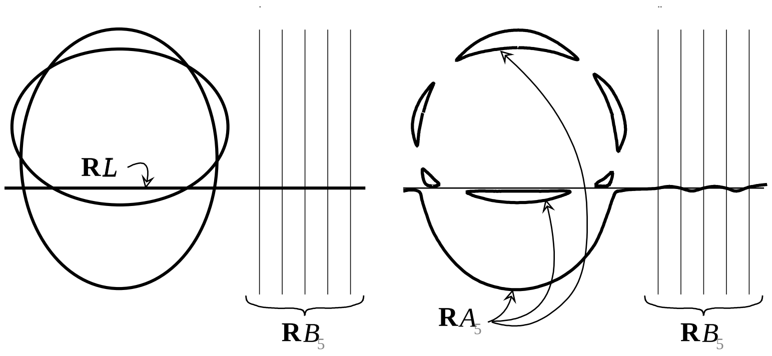

FIGURE 6

Proof. We shall actually construct a sequence of M-curves. At each step of the construction we add a line to the M-curve just constructed, and then give a slight perturbation to the union. We can begin the construction with a line or, as in Harnack's proof in [Har-76], with a circle. However, since we have already treated curves of degree ≤ 5 and constructed M-curves for those degrees (see Section 1.4), we shall begin by taking the M-curve of degree 5 that was constructed in Section 1.4, so that we can immediately proceed to curves that we have not encountered before.

Recall that we obtained a degree 5 M-curve by perturbing the union of two conics and a line L. This perturbation can be done using various curves. For what follows it is essential that the auxiliary curve intersect L in five points which are outside the two conics. For example, let the auxiliary curve be a union of five lines which satisfies this condition (Figure 6). We let B_5 denote this union, and we let A_5 denote the M-curve of degree 5 that is obtained using B_5.

We now construct a sequence of auxiliary curves B_m for $m > 5$. We take B_m to be a union of m lines which intersect L in m distinct points lying, for even m, in an arbitrary component of the set $\mathbb{R}L \smallsetminus \mathbb{R}B_{m-1}$ and for odd m in the component of $\mathbb{R}L \smallsetminus \mathbb{R}B_{m-1}$ containing $\mathbb{R}L \cap \mathbb{R}B_{m-2}$.

We construct the M-curve A_m of degree m using small perturbation of the union $A_{m-1} \cup L$ directed to B_m. Suppose that the M-curve A_{m-1} of degree $m-1$ has already been constructed, and suppose that $\mathbb{R}A_{m-1}$ intersects $\mathbb{R}L$ transversally in the $m-1$ points of the intersection $\mathbb{R}L \cap \mathbb{R}B_{m-1}$ which lie in the same component of the curve $\mathbb{R}A_{m-1}$ and in the same order as on $\mathbb{R}L$. It is not hard to see that, for one of the two possible directions of a small perturbation of $A_{m-1} \cup L$ directed to B_m, the line $\mathbb{R}L$ and the component of $\mathbb{R}A_{m-1}$ that it intersects give

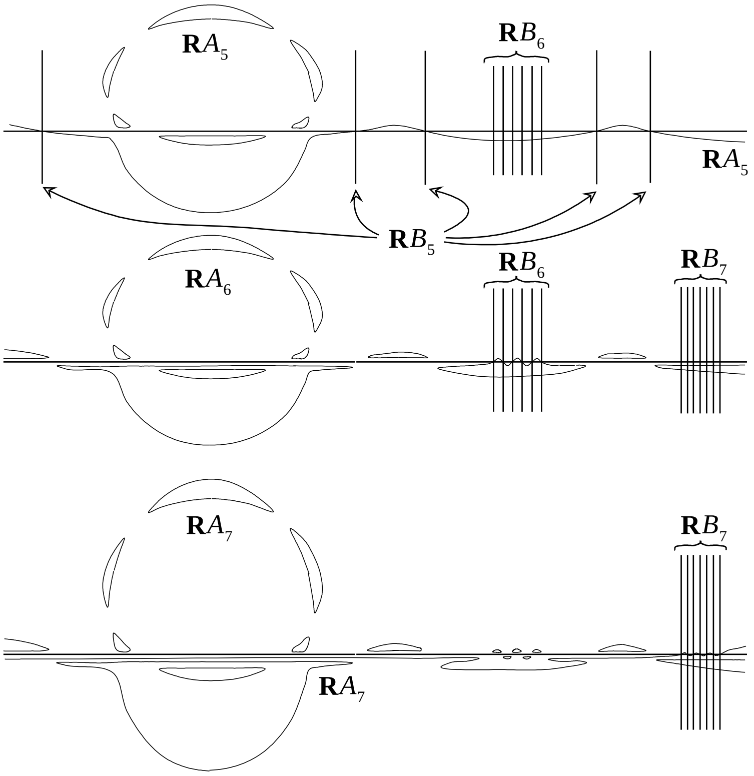

FIGURE 7

$m-1$ components, while the other components of $\mathbb{R}A_{m-1}$, of which, by assumption, there are

$$((m-1)^2 - 3(m-1) + 4)/2 - 1 = (m^2 - 5m + 6)/2,$$

are only slightly deformed—so that the number of components of $\mathbb{R}A_m$ remains equal to $(m^2 - 5m + 6)/2 + m - 1 = (m^2 - 3m + 4)/2$. We have thus obtained an M-curve of degree m. This curve is transversal to $\mathbb{R}L$, it intersects $\mathbb{R}L$ in $\mathbb{R}L \cap \mathbb{R}B_m$ (see *1.5.A*), and, since $\mathbb{R}L \cap \mathbb{R}B_m$ is contained in one of the components of the set $\mathbb{R}L \smallsetminus \mathbb{R}B_{m-1}$, it follows that the intersection points of our curve with $\mathbb{R}L$ are all in the same component of the curve and are in the same order as on $\mathbb{R}L$ (Figure 7). $\square$

The proof that the left inequality in 1 is best possible, i.e., that there is a curve with the minimum number of components, is much simpler. For example, we can take the curve given by the equation $x_0^m + x_1^m + x_2^m = 0$. Its set of real points is obviously empty when m is even, and when m is odd the set of real points is homeomorphic to

$\mathbb{R}P^1$ (we can get such a homeomorphism onto $\mathbb{R}P^1$, for example, by projection from the point $(0:0:1)$).

By choosing the auxiliary curves B_m in different ways in the construction of M-curves in the proof of Theorem *1.6.B*, we can obtain curves with any intermediate number of components. However, to complete the proof of Theorem *1.6.A* in this way would be rather tedious, even though it would not require any new ideas. We shall instead turn to a less explicit, but simpler and more conceptual method of proof, which is based on objects and phenomena not encountered above.

1.7. **Digression: the Space of Real Projective Plane Curves.** By the definition of real projective algebraic plane curves of degree m, they form a real projective space of dimension $m(m+3)/2$. The homogeneous coordinates in this projective space are the coefficients of the polynomials defining the curves. We shall denote this space by the symbol $\mathbb{R}C_m$. Its only difference with the standard space $\mathbb{R}P^{m(m+3)/2}$ is the unusual numbering of the homogeneous coordinates. The point is that the coefficients of a homogeneous polynomial in three variables have a natural double indexing by the exponents of the monomials:

$$a(x_0,x_1,x_2)=\sum\nolimits_{\substack{i,j\ge 0\\ i+j\le m}} a_{ij}x_0^{m-i-j}x_1^ix_2^j.$$

We let $\mathbb{R}NC_m$ denote the subset of $\mathbb{R}C_m$ corresponding to the real nonsingular curves. It is obviously open in $\mathbb{R}C_m$. Moreover, any nonsingular curve of degree m has a neighborhood in $\mathbb{R}NC_m$ consisting of isotopic nonsingular curves. Namely, small changes in the coefficients of the polynomial defining the curve lead to polynomials which give smooth sections of a tubular fibration of the original curve. This is an easy consequence of the implicit function theorem; compare with *1.5.A*, condition (3).

Curves which belong to the same component of the space $\mathbb{R}NC_m$ of nonsingular degree m curves are isotopic—this follows from the fact that nonsingular curves which are close to one another are isotopic. A path in $\mathbb{R}NC_m$ defines an isotopy in $\mathbb{R}P^2$ of the set of real points of a curve. An isotopy obained in this way is made of sets of real points of of real points of curves of degree m. Such an isotopy is said to be *rigid*. This definition naturally gives rise to the following classification problem, which is every bit as classical as problems *1.1.A*and *1.1.B*.

1.7.A (Rigid Isotopy Classification Problem)**.** *Classify the nonsingular curves of degree m up to rigid isotopy, i.e., study the partition of the space $\mathbb{R}NC_m$ of nonsingular degree m curves into its components.*

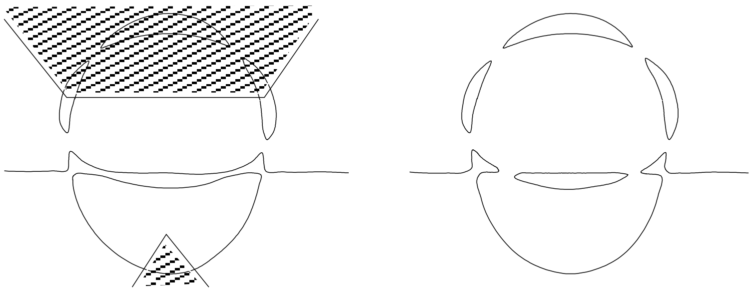

FIGURE 8

If $m \leq 2$, it is well known that the solution of this problem is identical to that of problem *1.1.B*. Isotopy also implies rigid isotopy for curves of degree 3 and 4. This was known in the last century; however, we shall not discuss this further here, since it has little relevance to what follows. At present problem *1.7.A* has been solved for $m \leq 6$.

Although this section is devoted to the early stages of the theory, I cannot resist commenting in some detail about a more recent result. In 1978, V. A. Rokhlin [Rok-78] discovered that for $m \geq 5$ isotopy of nonsingular curves of degree m no longer implies rigid isotopy. The simplest example is given in Figure 8, which shows two curves of degree 5. They are obtained by slightly perturbing the very same curve in Figure 4 which is made up of two conics and a line. Rokhlin's original proof uses argument on complexification, it will be presented below, in Section ??? Here, to prove that these curves are not rigid isotopic, we use more elementary arguements. Note that the first curve has an oval lying inside a triangle which does not intersect the one-sided component and which has its vertices inside the other three ovals, and the second curve does not have such an oval—but under a rigid isotopy the oval cannot leave the triangle, since that would entail a violation of Bézout's theorem.

We now examine the subset of $\mathbb{R}C_m$ made up of real singular curves.

It is clear that a curve of degree m has a singularity at $(1 : 0 : 0)$ if and only if its polynomial has zero coefficients of the monomials $x_0^m, x_0^{m-1}x_1, x_0^{m-1}x_2$. Thus, the set of real projective plane curves of degree m having a singularity at a particular point forms a subspace of codimension 3 in $\mathbb{R}C_m$.

We now consider the space S of pairs of the form (p, C), where $p \in \mathbb{R}P^2$, $C \in \mathbb{R}C_m$, and p is a singular point of the curve C. S is clearly an algebraic subvariety of the product $\mathbb{R}P^2 \times \mathbb{R}C_m$. The restriction to S of the projection $\mathbb{R}P^2 \times \mathbb{R}C_m \to \mathbb{R}P^2$ is a locally trivial fibration whose fiber is the space of curves of degree m with a singularity at the corresponding point, i.e., the fiber is a projective space of dimension $m(m+3)/2-3$. Thus, S is a smooth manifold of dimension $m(m+3)/2-1$. The restriction $S \to \mathbb{R}C_m$ of the projection $\mathbb{R}P^2 \times \mathbb{R}C_m \to \mathbb{R}C_m$ has as its image precisely the set of all real singular curves of degree m, i.e., $\mathbb{R}C_m \smallsetminus \mathbb{R}NC_m$. We let $\mathbb{R}SC_m$ denote this image. Since it is the image of a $(m(m+3)/2-1)$-dimensional manifold under smooth map, its dimension is at most $m(m+3)/2-1$. On the other hand, its dimension is at least equal $m(m+3)/2-1$, since otherwise, as a subspace of codimension ≥ 2, it would not separate the space $\mathbb{R}C_m$, and all nonsingular curves of degree m would be isotopic.

Using an argument similar to the proof that $\dim \mathbb{R}SC_m \leq m(m+3)/2-1$, one can show that the set of curves having at least two singular points and the set of curves having a singular point where the matrix of second derivatives of the corresponding polynomial has rank ≤ 1, each has dimension at most $m(m+3)/2-2$. Thus, the set $\mathbb{R}SC_m$ has an open everywhere dense subset consisting of curves with only one singular point, which is a nondegenerate double point (meaning that at this point the matrix of second derivatives of the polynomial defining the curve has rank 2). This subset is called the *principal part* of the set $\mathbb{R}SC_m$. It is a smooth submanifold of codimension 1 in $\mathbb{R}C_m$. In fact, its preimage under the natural map $S \to \mathbb{R}C_m$ is obviously an open everywhere dense subset in the manifold S, and the restriction of this map to the preimage is easily verified to be a one-to-one immersion, and even a smooth imbedding.

There are two types of nondegenerate real points on a plane curve. We say that a nondegenerate real double point $(\xi_0 : \xi_1 : \xi_2)$ on a curve A is *solitary* if the matrix of second partial derivatives of the polynomial defining A has either two nonnegative or two nonpositive eigenvalues at the point (ξ_0, ξ_1, ξ_2). A solitary nondegenerate double point of A is an isolated point of the set $\mathbb{R}A$. In general, a singular point of A which is an isolated point of the set $\mathbb{R}A$ will be called a solitary real singular point. The other type of nondegenerate real double point is a crossing; crossings were discussed in Section 1.5 above. Corresponding to this division of the nondegenerate real double points into solitary points and crossings, we have a partition of the principal part of the set of real singular curves of degree m into two open sets.

If a curve of degree m moves as a point of $\mathbb{R}C_m$ along an arc which has a transversal intersection with the half of the principal part of the set of real singular curves consisting of curves with a solitary singular point, then the set of real points on this curve undergoes a Morse modification of index 0 or 2 (i.e., either the curve acquires a solitary double point, which then becomes a new oval, or else one of the ovals contracts to a point (a solitary nondegenerate double point) and disappears). In the case of a transversal intersection with the other half of the principal part of the set of real singular curves one has a Morse modification of index 1 (i.e., two arcs of the curve approach one another and merge, with a crossing at the point where they come together, and then immediately diverge in their modified form, as happens, for example, with the hyperbola in the family of affine curves of degree 2 given by the equation $xy = t$ at the moment when $t = 0$).

A line in $\mathbb{R}C_m$ is called a (real) pencil of curves of degree m. If a and b are polynomials defining two curves of the pencil, then the other curves of the pencil are given by polynomials of the form $\lambda a + \mu b$ with $\lambda, \mu \in \mathbb{R} \smallsetminus 0$.

By the transversality theorem, the pencils which intersect the set of real singular curves only at points of the principal part and only transversally form an open everywhere dense subset of the set of all real pencils of curves of degree m.

1.8. **End of the Proof of Theorem *1.6.A*.** In Section 1.6 it was shown that for any m there exist nonsingular curves of degree m with the minimum number $(1 - (-1)^m)/2$ or with the maximum number $(m^2 - 3m + 4)/2$ of components. Nonsingular curves which are isotopic to one another form an open set in the space $\mathbb{R}C_m$ of real projective plane curves of degree m (see Section 1.7). Hence, there exists a real pencil of curves of degree m which connects a curve with minimum number of components to a curve with maximum number of components and which intersects the set of real singular curves only in its principal part and only transversally. As we move along this pencil from the curve with minimum number of components to the curve with maximum number of components, the curve only undergoes Morse modifications, each of which changes the number of components by at most 1. Consequently, this pencil includes nonsingular curves with an arbitrary intermediate number of components. □

1.9. **Isotopy Types of Harnack M-Curves.** Harnack's construction of M-curves in [Har-76] differs from the construction in the proof of Theorem *1.6.B* in that a conic, rather than a curve of degree 5, is used as the original curve. Figure 9 shows that the M-curves of degree ≤ 5 which are used in Harnack's construction [Har-76]. For $m \geq 6$ Harnack's construction gives M-curves with the same isotopy types as in the construction in Section 1.6.

In these constructions one obtains different isotopy types of M-curves depending on the choice of auxiliary curves (more precisely, depending on the relative location of the intersections $\mathbb{R}B_m \cap \mathbb{R}L$). Recall that in

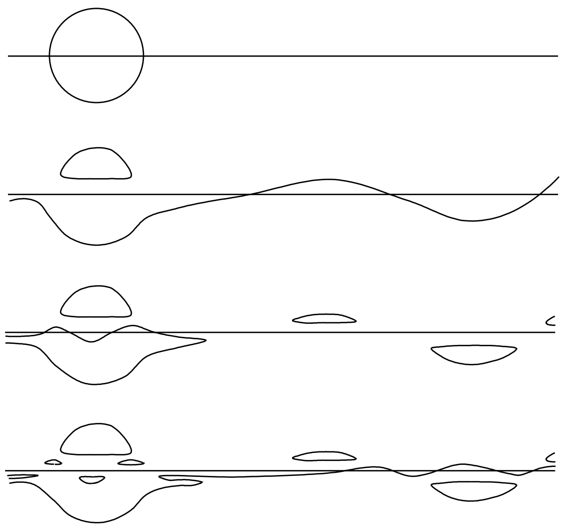

FIGURE 9

order to obtain M-curves it is necessary for the intersection $\mathbb{R}B_m \cap \mathbb{R}L$ to consist of m points and lie in a single component of the set $\mathbb{R}L \smallsetminus \mathbb{R}B_{m-1}$, where for odd m this component must contain $\mathbb{R}B_{m-2} \cap \mathbb{R}L$. It is easy to see that the isotopy type of the resulting M-curve of degree m depends only on the choice of the components of $\mathbb{R}L \smallsetminus \mathbb{R}B_{r-1}$ for even $r < m$ where the intersections $\mathbb{R}L \cap \mathbb{R}B_r$ are to be found. If we take the components containing $\mathbb{R}L \cap \mathbb{R}B_{r-2}$ for even r as well, then the degree m M-curve obtained from the construction has isotopy type $\langle J \amalg (m^2-3m+2)/2 \rangle$ for odd m and $\langle (3m^2-6m)/8 \amalg 1 \langle (m^2-6m+8)/8 \rangle \rangle$ for even m. In Table 2 we have listed the isotopy types of M-curves of degree ≤ 10 which one obtains from Harnack's construction using all possible B_m.

In conclusion, we mention two curious properties of Harnack M-curves, for which the reader can easily furnish a proof.

1.9.A. *The depth of a nest in a Harnack M-curve is at most 2.*

1.9.B. *Any Harnack M-curve of even degree m has $(3m^2 - 6m + 8)/8$ outer ovals and $(m^2 - 6m + 8)/8$ inner ovals.*

1.10. **Hilbert Curves.** In 1891 Hilbert [Hil-91] seems to have been the first to clearly state the isotopy classification problem for nonsingular curves. As we saw, the isotopy types of Harnack M-curves are very special. Hilbert suggested that from the topological viewpoints

TABLE 2

m	Isotopy types of the Harnack M-curves of degree m			
2	$\langle 1\rangle$			
3	$\langle J \amalg 1\rangle$			
4	$\langle 4\rangle$			
5	$\langle J \amalg 6\rangle$			
6	$\langle 9 \amalg 1\langle 1\rangle\rangle$			
7	$\langle J \amalg 15\rangle$	$\langle J \amalg 13 \amalg 1\langle 1\rangle\rangle$		
8	$\langle 18 \amalg 1\langle 3\rangle\rangle$	$\langle 17 \amalg 1\langle 1\rangle \amalg 1\langle 2\rangle\rangle$		
9	$\langle J \amalg 28\rangle$	$\langle J \amalg 24 \amalg 1\langle 3\rangle\rangle$	$\langle J \amalg 26 \amalg 1\langle 1\rangle\rangle$	$\langle J \amalg 23 \amalg 1\langle 1\rangle \amalg 1\langle 2\rangle\rangle$
10	$\langle 30 \amalg 1\langle 6\rangle\rangle$	$\langle 29 \amalg 2\langle 3\rangle\rangle$	$\langle 29 \amalg 1\langle 1\rangle \amalg 1\langle 5\rangle\rangle$	$\langle 28 \amalg 1\langle 1\rangle \amalg 1\langle 2\rangle \amalg 1\langle 3\rangle\rangle$

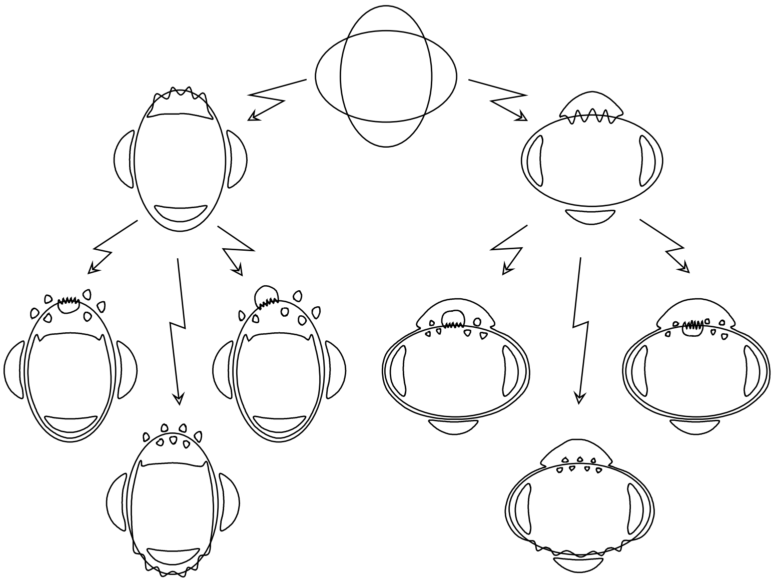

FIGURE 10. Construction of even degree curves by Hilbert's method. Degrees 4 and 6.

M-curves are the most interesting. This Hilbert's guess was strongly confirmed by the whole subsequent development of the field.

There is a big gap between property *1.9.A* of Harnack M-curves and the corresponding prohibition in *1.3.C*. Hilbert [Hil-91] showed that this gap is explained by the peculiarities of the construction and not by the intrinsic properties of M-curves. He proposed a new method of

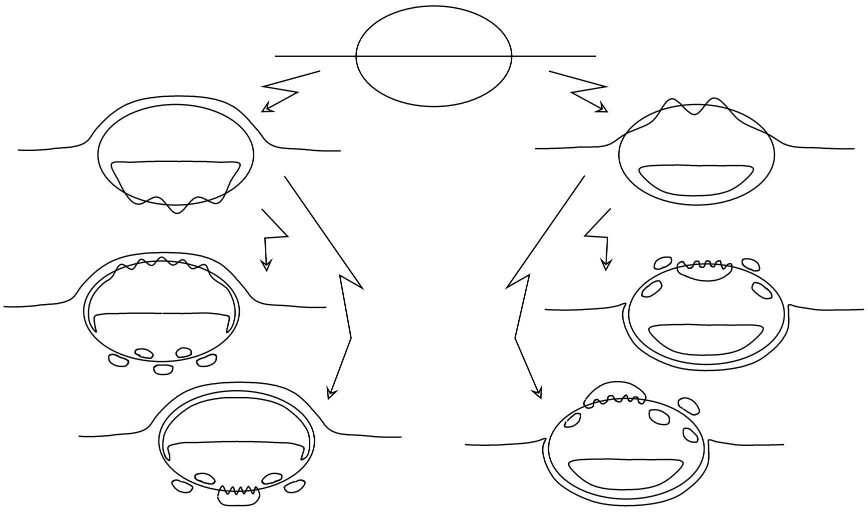

Figure 11. Construction of odd degree curves by Hilbert's method. Degrees 3 and 5.

Table 3

m	Isotopy types of the Hilbert M-curves of degree m
4	$\langle 1\rangle$
6	$\langle 9 \amalg 1\langle 1\rangle\rangle$ $\langle 1 \amalg 1\langle 9\rangle\rangle$
8	$\langle 5 \amalg 1\langle 14 \amalg 1\langle 1\rangle\rangle\rangle$ $\langle 17 \amalg 1\langle 2 \amalg 1\langle 1\rangle\rangle\rangle$ $\langle 18 \amalg 1\langle 3\rangle\rangle$ $\langle 1 \amalg 1\langle 2 \amalg 1\langle 17\rangle\rangle\rangle$ $\langle 1 \amalg 1\langle 14 \amalg 1\langle 5\rangle\rangle\rangle$
5	$\langle J \amalg 6\rangle$
7	$\langle J \amalg 15\rangle$ $\langle J \amalg 12 \amalg 1\langle 2\rangle\rangle$ $\langle J \amalg 13 \amalg 1\langle 3\rangle\rangle$ $\langle J \amalg 2 \amalg 1\langle 12\rangle\rangle$ $\langle J \amalg 1 \amalg 1\langle 13\rangle\rangle$

constructing M-curves which was close to Harnack's method, but which gives M-curves with nests of any depth allowed by Theorem *1.3.C*. In his method the role a line plays in Harnack's method is played instead by a nonsingular conic, and a line or a conic is used for the starting curve. Figures 10–11 show how to construct M-curves by Hilbert's method.

In Table 3 we list the isotopy types of M-curves of degree ≤ 8 which are obtained by Hilbert's construction.

The first difficult special problems that Hilbert met were related with curves of degree 6. Hilbert succeeded to construct M-curves of degree ≥ 6 with mutual position of components different from the scheme $\langle 9 \amalg 1\langle 1\rangle\rangle$ realized by Harnack. However he realized only one new real scheme of degree 6, namely $\langle 1 \amalg 1\langle 9\rangle\rangle$. Hilbert conjectured that these are the only real schemes realizable by M-curves of degree 6 and for a long time affirmed that he had a (long) proof of this conjecture. Even being false (it was disproved by D. A. Gudkov in 1969, who constructed a curve with the scheme $\langle 5 \amalg 1\langle 5\rangle\rangle$) this conjecture caught the things that became in 30-th and 70-th the core of the theory.

In fact, Hilbert invented a method which allows to answer to all questions on topology of curves of degree 6. It involves a detailed analysis of singular curves which could be obtained from a given nonsingular one. The method required complicated fragments of singularity theory, which had not been elaborated at the time of Hilbert. Completely this project was realized only in the sixties by D. A. Gudkov. It was Gudkov who obtained a complete table of real schemes of curves of degree 6.

Coming back to Hilbert, we have to mention his famous problem list [Hil-01]. He included into the list, as a part of the sixteenth problem, a general question on topology of real algebraic varieties and more special questions like the problem on mutual position of components of a plane curve of degree 6.

The most mysterious in this problem seems to be its number. The number *sixteen* plays a very special role in topology of real algebraic varieties. It is difficult to believe that Hilbert was aware of that. It became clear only in the beginning of seventies (see Rokhlin's paper "Congruences modulo sixteen in the sixteenth Hilbert's problem" [Rok-72]). Nonetheless, sixteen was the number assigned by Hilbert to the problem.

1.11. **Analysis of the Results of the Constructions. Ragsdale.** In 1906, V. Ragsdale [Rag-06] made a remarkable attempt to guess new prohibitions, based on the results of the constructions by Harnack's and Hilbert's methods. She concentrated her attention on the case of curves of even degree, motivated by the following special properties of such curves. Since a curve of an even degree is two-sided, it divides $\mathbb{R}P^2$ into two parts, which have the curve as their common boundary. One of the parts contains a nonorientable component; it is denoted by $\mathbb{R}P^2_-$. The other part, which is orientable, is denoted by $\mathbb{R}P^2_+$. The ovals of a curve of even degree are divided into inner and outer ovals with respect to $\mathbb{R}P^2_+$ (i.e., into ovals which bound a component of $\mathbb{R}P^2_+$ from the inside and from the outside). Following Petrovsky [Pet-38], one says that the outer ovals with respect to $\mathbb{R}P^2_+$ are the *even* ovals (since such an oval lies inside an even number of other ovals), and the rest of the ovals are called *odd* ovals. The number of even ovals is denoted by p, and the number of odd ovals is denoted by n. These numbers contain very important information about the topology of the sets $\mathbb{R}P^2_+$ and $\mathbb{R}P^2_-$. Namely, the set $\mathbb{R}P^2_+$ has p components, the set $\mathbb{R}P^2_-$ has $n+1$ components, and the Euler characteristics are given by $\chi(\mathbb{R}P^2_+) = p-n$ and $\chi(\mathbb{R}P^2_-) = n-p+1$. Hence, one should pay special attention to the numbers p and n. (It is amazing that essentially these considerations were stated in a paper in 1906!)

By analyzing the constructions, Ragsdale [Rag-06] made the following observations.

1.11.A ((compare with *1.9.A* and *1.9.B*)). *For any Harnack M-curve of even degree m,*

$$p = (3m^2 - 6m + 8)/8, \qquad n = (m^2 - 6m + 8)/8.$$

1.11.B. *For any Hilbert M-curve of even degree m,*

$$(m^2 - 6m + 16)/8 \le p \le (3m^2 - 6m + 8)/8,$$
$$(m^2 - 6m + 8)/8 \le n \le (3m^2 - 6m)/8.$$

This gave her evidence for the following conjecture.

1.11.C (Ragsdale Conjecture). *For any curve of even degree m,*

$$p \le (3m^2 - 6m + 8)/8, \qquad n \le (3m^2 - 6m)/8.$$

The most mysterious in this problem seems to be its number. The number *sixteen* plays a very special role in topology of real algebraic varieties. It is difficult to believe that Hilbert was aware of that. It became clear only in the beginning of seventies (see Rokhlin's paper "Congruences modulo sixteen in the sixteenth Hilbert's problem" [Rok-72]). Nonetheless, sixteen was the number assigned by Hilbert to the problem.

Writing cautiously, Ragsdale formulated also weaker conjectures. About thirty years later I. G. Petrovsky [Pet-33], [Pet-38] proved one of these weaker conjectures. See below Subsection 1.13.

Petrovsky also formulated conjectures about the upper bounds for p and n. His conjecture about n was more cautious (by 1).

Both Ragsdale Conjecture formulated above and its version stated by Petrovsky [Pet-38] are wrong. However they stayed for rather long time: Ragsdale Conjecture on n was disproved by the author of this book [Vir-80] in 1979. However the disproof looked rather like improvement of the conjecture, since in the counterexamples $n = \frac{3k(k-1)}{2} + 1$. Drastically Ragsdale-Petrovsky bounds were disproved by I. V. Itenberg [Ite-93] in 1993: in Itenberg's counterexamples the difference between p (or n) and $\frac{3k(k-1)}{2} + 1$ is a quadratic function of k.

In Section **??** we shall return to this very first conjecture of a general nature on the topology of real algebraic curves. At this point we shall only mention that several weaker assertions have been proved and examples have been constructed which made it necessary to weaken the second inequality by 1. In the weaker form the Ragsdale conjecture has not yet been either proved or disproved.

The numbers p and n introduced by Ragsdale occur in many of of the prohibitions that were subsequently discovered. While giving full credit to Ragsdale for her insight, we must also say that, if she had looked more carefully at the experimental data available to her, she should have been able to find some of these prohibitions. For example, it is not clear what stopped her from making the conjecture which was made by Gudkov [GU-69] in the late 1960's. In particular, the experimental data could suggest the formulation of the Gudkov-Rokhlin congruence proved in [Rok-72]: *for any M-curve of even degree $m = 2k$*

$$p - n \equiv k^2 \mod 8$$

Maybe mathematicians trying to conjecture restrictions on some integer should keep this case in mind as an evidence that restrictions can have not only the shape of inequality, but congruence. Proof of these Gudkov's conjectures initiated by Arnold [Arn-71] and completed by Rokhlin [Rok-72], Kharlamov [Kha-73], Gudkov and Krakhnov [GK-73] had marked the beginning of the most recent stage in the development of the topology of real algebraic curves. We shall come to this story at the end of this Section.

1.12. **Generalizations of Harnack's and Hilbert's Methods. Brusotti. Wiman.** Ragsdale's work [Rag-06] was partly inspired by the erroneous paper of Hulbrut, containing a proof of the false assertion that an M-curve can be obtained by means of a classical small perturbation (see Section 1.5) from only two M-curves, one of which must have degree ≤ 2. If this had been true, it would have meant that an inductive construction of M-curves by classical small perturbations starting with curves of small degree must essentially be either Harnack's method or Hilbert's method.

In 1910–1917, L. Brusotti showed that this is not the case. He found inductive constructions of M-curves based on classical small perturbation which were different from the methods of Harnack and Hilbert.

Before describing Brusotti's constructions, we need some definitions. A simple arc X in the set of real points of a curve A of degree m is said to be a *base of rank* ρ if there exists a curve of degree ρ which intersects the arc in ρm (distinct) points. A base of rank ρ is clearly also a base of rank any multiple of ρ (for example, one can obtain the intersecting curve of the corresponding degree as the union of several copies of the degree ρ curve, each copy shifted slightly).

An M-curve A is called a *generating curve* if it has disjoint bases X and Y whose ranks divide twice the degree of the curve. An M-curve A_0 of degree m_0 is called an *auxiliary curve* for the generating curve A of degree m with bases X and Y if the following conditions hold:

a) The intersection $\mathbb{R}A \cap \mathbb{R}A_0$ consist of mm_0 distinct points and lies in a single component K of $\mathbb{R}A$ and in a single component K of $\mathbb{R}A_0$.

b) The cyclic orders determined on the intersection $\mathbb{R}A \cap \mathbb{R}A_0$ by how it is situated in K and in K_0 are the same.

c) $X \subset \mathbb{R}A \smallsetminus \mathbb{R}A_0$.

d) If K is a one-sided curve and $m_0 \equiv \mod 2$, then the base X lies outside the oval K_0.

e) The rank of the base X is a divisor of the numbers $m + m_0$ and $2m$, and the rank of Y is a divisor of $2m + m_0$ and $2m$.

An auxiliary curve can be the empty curve of degree 0. In this case the rank of X must be a divisor of the degree of the generating curve.

Let A be a generating curve of degree m, and let A_0 be a curve of degree m_0 which is an auxiliary curve with respect to A and the bases X and Y. Since the rank of X divides $m + m_0$, we may assume that the rank is equal to $m + m_0$. Let C be a real curve of degree $m + m_0$ which intersects X in $m(m + m_0)$ distinct points. It is not hard to verify that a classical small perturbation of the curve $A \cup A_0$ directed to L will give an M-curve of degree $m + m_0$, and that this M-curve will be an auxiliary curve with respect to A and the bases obtained from Y and X (the bases must change places). We can now repeat this construction, with A_0 replaced by the curve that has just been constructed. Proceeding in this way, we obtain a sequence of M-curves whose degree forms an arithmetic progression: $km + m_0$ with $k = 1, 2, \dots$. This is called the construction by Brusotti's method, and the sequence of M-curves is called a *Brusotti series*.

Any simple arc of a curve of degree ≤ 2 is a base of rank 1 (and hence of any rank). This is no longer the case for curves of degree ≤ 3. For example, an arc of a curve of degree 3 is a base of rank 1 if and only if it contains a point of inflection. (We note that a base of rank 2 on a curve of degree 3 might not contain a point of inflection: it might be on the oval rather than on the one-sided component where all of the points of inflection obviously lie. A curve of degree 3 with this type of base of rank 2 can be constructed by a classical small perturbation of a union of three lines.)

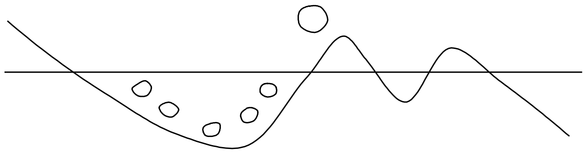

FIGURE 12

If the generating curve has degree 1 and the auxiliary curve has degree 2, then the Brusotti construction turns out to be Harnack's construction. The same happens if we take an auxiliary curve of degree 1 or 0. If the generating curve has degree 2 and the auxiliary curve has degree 1 or 2 (or 0), then the Brusotti construction is the same as Hilbert's construction.

In general, not all Harnack and Hilbert constructions are included in Brusotti's scheme; however, the Brusotti construction can easily be extended in such a way as to be a true generalization of the Harnack and Hilbert constructions. This extension involves allowing the use of an arbitrary number of bases of the generating curve. Such an extension is particularly worthwhile when the generating curve has degree ≤ 2, in which case there are arbitrarily many bases.

It can be shown that Brusotti's construction with generating curve of degree 1 and auxiliary curve of degree ≤ 4 gives the same types of M-curves as Harnack's construction. But as soon as one uses auxiliary curves of degree 5, one can obtain new isotopy types from Brusotti's construction. It was only in 1971 that Gudkov [Gud-71] found an auxiliary curve of degree 5 that did this. His construction was rather complicated, and so I shall only give some references [Gud-71], [Gud-74], [A'C-79] and present Figure 12, which illustrates the location of the degree 5 curve relative to the generating line.

Even with the first stage of Brusotti's construction, i.e., the classical small perturbation of the union of the curve and the line, one obtains an M-curve (of degree 6) which has isotopy type $\langle 5 \amalg 1\langle 5\rangle\rangle$, an isotopy type not obtained using the constructions of Harnack and Hilbert. Such an M-curve of degree 6 was first constructed in a much more complicated way by Gudkov [GU-69], [Gud-73] in the late 1960's.

In Figures 13 and 14 we show the construction of two curves of degree 6 which are auxiliary curves with respect to a line. In this case the Brusotti construction gives new isotopy types beginning with degree 8.

In the Hilbert construction we keep track of the location relative to a fixed line A. The union of two conics is perturbed in direction to a

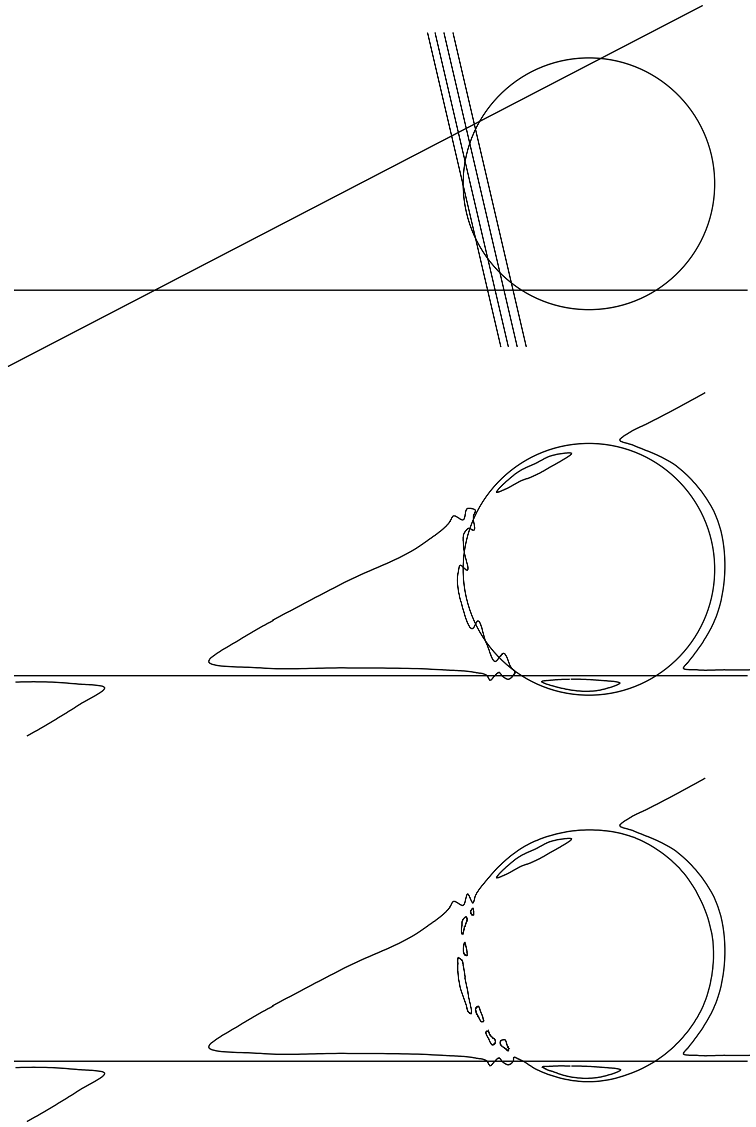

FIGURE 13

quadruple of lines. One obtains a curve of degree 4. To this curve one then adds one of the original conics, and the union is perturbed.

In numerous papers by Brusotti and his students, many series of Brusotti M-curves were found. Generally, new isotopy types appear in them beginning with degree 9 or 10. In these constructions they paid much attention to combinations of nests of different depths—a theme which no longer seems to be very interesting. An idea of the nature of the results in these papers can be obtained from Gudkov's

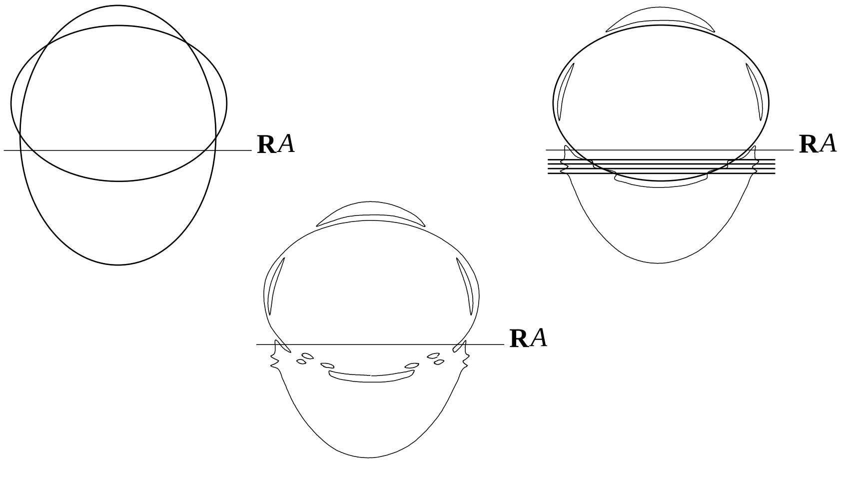

FIGURE 14. In the construction by Hilbert's method, we keep track of the locations relative to a fixed line A. The union of two conics is perturbed in direction to a 4-tuple of lines. A curve of degree 4 is obtained. We add one of the original conics to this curve, and then perturb the union.

survey [Gud-74]; for more details, see Brusotti's survey [Bru-56] and the papers cited there.

An important variant of the classical constructions of M-curves, of which we shall need to make use in the next section, is not subsumed under Brusotti's scheme even in its extended form. This variant, proposed by Wiman [Wim-23], consists in the following. We take an M-curve A of degree k having base X of rank dividing k; near this curve we construct a curve A' transversally intersecting A in k^2 points of X, after which we can subject the union $A \cup A'$ to a classical small perturbation, giving an M-curve of degree $2k$ (for example, a perturbation in direction to an empty curve of degree $2k$). The resulting M-curve has the following topological structure: each of the components of the curve A except for one (i.e., except for the component containing X) is doubled, i.e., is replaced by a pair of ovals which are each close to an oval of the original curve, and the component containing X gives a chain of k^2 ovals. This new curve does not necessarily have a base, so that in general one cannot construct a series of M-curves in this way.

1.13. **The First Prohibitions not Obtained from Bézout's Theorem.** The techniques discussed above are, in essence, completely elementary. As we saw (Section 1.4), they are sufficient to solve the isotopy classification problem for nonsingular projective curves of degree ≤ 5. However, even in the case of curves of degree 6 one needs subtler considerations. Not all of the failed attempts to construct new isotopy types of M-curves of degree 6 (after Hilbert's 1891 paper [Hil-91], there were two that had not been realized: $\langle 9 \amalg 1\langle 1\rangle\rangle$ and $\langle 1 \amalg 1\langle 9\rangle\rangle$) could be explained on the basis of Bézout's theorem. Hilbert undertook an attack on M-curves of degree 6. He was able to grope his way toward a proof that isotopy types cannot be realized by curves of degree 6, but the proof required a very involved investigation of the natural stratification of the space $\mathbb{R}C_6$ of real curves of degree 6. In [Roh-13], Rohn, developing Hilbert's approach, proved (while stating without proof several valid technical claims which he needed) that the types $\langle 11\rangle$ and $\langle 1\langle 10\rangle\rangle$ cannot be realized by curves of degree 6. It was not until the 1960's that the potential of this approach was fully developed by Gudkov. By going directly from Rohn's 1913 paper [Roh-13] to the work of Gudkov, I would violate the chronological order of my presentation of the history of prohibitions. But in fact I would only be omitting one important episode, to be sure a very remarkable one: the famous work of I. G. Petrovsky [Pet-33], [Pet-38] in which he proved the first prohibition relating to curves of arbitrary even degree and not a direct consequence of Bézout's theorem.

1.13.A (Petrovsky Theorem ([Pet-33], [Pet-38]))**.** *For any nonsingular real projective algebraic plane curve of degree $m = 2k$*

$$-\frac{3}{2}k(k-1) \le p - n \le \frac{3}{2}k(k-1) + 1. \tag{2}$$

(Recall that p denotes the number of even ovals on the curve (i.e., ovals each of which is enveloped by an even number of other ovals, see Section 1.11), and n denotes the number of odd ovals.)

As it follows from [Pet-33] and [Pet-38], Petrovsky did not know Ragsdale's paper. But his proof runs along the lines indicated by Ragsdale. He also reduced the problem to estimates of Euler characteristic of the pencil curves, but he went further: he proved these estimates. Petrovsky's proof was based on a technique that was new in the study of the topology of real curves: the Euler-Jacobi interpolation formula. Petrovsky's theorem was generalized by Petrovsky and Oleinik [PO-49] to the case of varieties of arbitrary dimension, and by Oleĭnik [Ole-51] to the case of curves on a surface. More about the proof and the influence of Petrovsky's work on the subsequent development of the subject can be found in Kharlamov's survey [Kha-86] in Petrovsky's collected works. I will only comment that in application to nonsingular projective plane curves, the full potential of Petrovsky's method, insofar as we are able to judge, was immediately realized by Petrovsky himself.

We now turn to Gudkov's work. In a series of papers in the 1950's and 1960's, he completed the development of the techniques needed to realize Hilbert's approach to the problem of classifying curves of degree 6 (these techniques were referred to as the Hilbert-Rohn method by Gudkov), and he used the techniques to solve this problem (see [GU-69]). The answer turned out to be elegant and stimulating.

1.13.B (Gudkov's Theorem [GU-69]). *The* 56 *isotopy types listed in Table 4, and no others, can be realized by nonsingular real projective algebraic plane curves of degree* 6*.*

$\langle 9 \amalg 1\langle 1\rangle\rangle$ $\quad$ $\langle 5 \amalg 1\langle 5\rangle\rangle$ $\quad$ $\langle 1 \amalg 1\langle 9\rangle\rangle$

$\langle 10\rangle$ $\langle 8 \amalg 1\langle 1\rangle\rangle$ $\quad$ $\langle 5 \amalg 1\langle 4\rangle\rangle$ $\langle 4 \amalg 1\langle 5\rangle\rangle$ $\quad$ $\langle 1 \amalg 1\langle 8\rangle\rangle$ $\langle 1\langle 9\rangle\rangle$

$\langle 9\rangle$ $\langle 7 \amalg 1\langle 1\rangle\rangle$ $\langle 6 \amalg 1\langle 2\rangle\rangle$ $\langle 5 \amalg 1\langle 3\rangle\rangle$ $\langle 4 \amalg 1\langle 4\rangle\rangle$ $\langle 3 \amalg 1\langle 5\rangle\rangle$ $\langle 2 \amalg 1\langle 6\rangle\rangle$ $\langle 1 \amalg 1\langle 7\rangle\rangle$ $\langle 1\langle 8\rangle\rangle$

$\langle 8\rangle$ $\langle 6 \amalg 1\langle 1\rangle\rangle$ $\langle 5 \amalg 1\langle 2\rangle\rangle$ $\langle 4 \amalg 1\langle 3\rangle\rangle$ $\langle 3 \amalg 1\langle 4\rangle\rangle$ $\langle 2 \amalg 1\langle 5\rangle\rangle$ $\langle 1 \amalg 1\langle 6\rangle\rangle$ $\langle 1\langle 7\rangle\rangle$

$\langle 7\rangle$ $\langle 5 \amalg 1\langle 1\rangle\rangle$ $\langle 4 \amalg 1\langle 2\rangle\rangle$ $\langle 3 \amalg 1\langle 3\rangle\rangle$ $\langle 2 \amalg 1\langle 4\rangle\rangle$ $\langle 1 \amalg 1\langle 5\rangle\rangle$ $\langle 1\langle 6\rangle\rangle$

$\langle 6\rangle$ $\langle 4 \amalg 1\langle 1\rangle\rangle$ $\langle 3 \amalg 1\langle 2\rangle\rangle$ $\langle 2 \amalg 1\langle 3\rangle\rangle$ $\langle 1 \amalg 1\langle 4\rangle\rangle$ $\langle 1\langle 5\rangle\rangle$

$\langle 5\rangle$ $\langle 3 \amalg 1\langle 1\rangle\rangle$ $\langle 2 \amalg 1\langle 2\rangle\rangle$ $\langle 1 \amalg 1\langle 3\rangle\rangle$ $\langle 1\langle 4\rangle\rangle$

$\langle 4\rangle$ $\langle 2 \amalg 1\langle 1\rangle\rangle$ $\langle 1 \amalg 1\langle 2\rangle\rangle$ $\langle 1\langle 3\rangle\rangle$

$\langle 3\rangle$ $\langle 1 \amalg 1\langle 1\rangle\rangle$ $\langle 1\langle 2\rangle\rangle$ $\quad$ $\langle 1\langle 1\langle 1\rangle\rangle\rangle$

$\langle 2\rangle$ $\langle 1\langle 1\rangle\rangle$

$\langle 1\rangle$

$\langle 0\rangle$

TABLE 4. Isotopy types of nonsingular real projective algebraic plane curves of degree 6.

This result, along with the available examples of curves of higher degree, led Gudkov to the following conjectures.

1.13.C (Gudkov Conjectures [GU-69]). (i) *For any M-curve of even degree $m = 2k$*

$$p - n \equiv k^2 \mod 8.$$

(ii) *For any $(M-1)$-curve of even degree $m = 2k$*

$$p - n \equiv k^2 \pm 1 \mod 8.$$

While attempting to prove conjecture *1.13.C*(i), V. I. Arnold [Arn-71] discovered some striking connections between the topology of a real algebraic plane curve and the topology of its complexification. Although he was able to prove the conjecture itself only in a weaker form (modulo 4 rather than 8), the new point of view he introduced to the subject opened up a remarkable perspective, and in fact immediately brought fruit: in the same paper [Arn-71] Arnold proved several new prohibitions (in particular, he strengthened Petrovsky's inequalities *1.13.A*). The full conjecture *1.13.C*(i) and its high-dimensional generalizations were proved by Rokhlin [Rok-72], based on the connections discovered by Arnold in [Arn-71].

I am recounting this story briefly here only to finish the preliminary history exposition. At this point the technique aspects are getting too complicated for a light exposition. After all, the prohibitions, which were the main contents of the development at the time we come to, are not the main subject of this book. Therefore I want to switch to more selective exposition emphasizing the most profound ideas rather than historical sequence of results.

A reader who prefare historic exposition can find it in Gudkov's survey article [Gud-74]. To learn about the many results obtained using methods from the modern topology of manifolds and complex algebraic geometry (the use of which was begun by Arnold in [Arn-71]), the reader is referred to the surveys [Wil-78], [Rok-78], [Arn-79], [Kha-78], [Kha-86], [Vir-86].

Exercises. 1.1 What is the maximal number p such that through any p points of $\mathbb{R}P^2$ one can trace a real algebraic curve of degree m?

1.2 Prove the Harnack inequality (the right hand side of (1)) deducing it from the Bézout Theorem.