...ifier elimination for elementary algebra and
...adjacency algorithm for cylindrical algebraic ... Comp. 5, (this issue).
...osed fields by cylindrical algebraic decompo... Automata Theory and Formal Languages.
...ACM SIGSAM Bull. 14, 19.
...nberger, B., Loos, R., Collins, G. E., eds.) ...utation (Computing Supplementum 4), pp.
...Proc. IEEE Conf. on Foundations
... (Buchberger, B., Loos, R., Collins, G. E., ...omputation (Computing Supplementum 4), ...g.
...m for cylindrical algebraic decomposition of ...ssue)
...aren Optimierung für beliebige Polynome mit ...erlag Anton Hain.
...ndrical algebraic decompositions. SIAM J.
...roblem II. General techniques for computing ...ate. Applied Math. 4, 298–351.

J. Symbolic Computation (1988) 5, 213–236

A Polynomial-time Algorithm for the Topological Type of a Real Algebraic Curve*

DENNIS S. ARNON AND SCOTT McCALLUM†

Xerox PARC, 3333 Coyote Hill Road, Palo Alto, California 94304, U.S.A.
† Research School of Physical Science, Australian National University, Canberra ACT 2601, Australia

(Received 13 February 1985, and in revised form 15 November 1987)

It was proved over a century ago that an algebraic curve C in the real projective plane, of degree n, has at most $\frac{(n-1)(n-2)}{2}+1$ connected components. If C is nonsingular, then each of its components is a topological circle. A circle in the projective plane either separates it into a disk (the *interior* of the circle) and a Möbius band (the circle's *exterior*), or does not separate it. In the former case, the circle is an *oval*. If C is nonsingular, then all its components are ovals if n is even, and all except one are ovals if n is odd. An oval is *included* in another if it lies in the other's interior. The topological type of (a nonsingular) C is completely determined by (1) the parity of n, (2) how many ovals it has, and (3) the partial ordering of its ovals by inclusion. We present an algorithm which, given a homogeneous polynomial $f(x, y, z)$ of degree n with integer coefficients, checks whether the curve defined by $f = 0$ is nonsingular, and if so, computes its topological type. The algorithm's maximum computing time is $O(n^{27}L(d)^3)$, where d is the sum of the absolute values of the integer coefficients of f, and $L(d)$ is the length of d.

1 Introduction

We begin with an example of what our algorithm does. Let $f(x, y, z)$ be the homogeneous polynomial

$$y^4 - 2xy^3 + x^2y^2 + y^2z^2 - 2x^3y + x^2z^2 - z^4 .$$

The equation $f = 0$ defines an algebraic curve C in the real projective plane. Let us draw a picture of C in two steps. Suppose that points in the projective plane have homogeneous xyz coordinates: then the points for which $z = 1$ constitute an affine xy-plane imbedded in the projective plane. The portion of C lying in this xy-plane is the locus of the equation

$$f(x, y, 1) = y^4 - 2xy^3 + x^2y^2 + y^2 - 2x^3y + x^2 - 1 = 0,$$

and is shown on the left in Fig. 1. Using the standard disk model for the projective plane, the full curve C is shown on the right in Fig. 1. C is of degree four, and happens

* This work was supported by the National Science Foundation (Grant MCS-8009357 to the University of Wisconsin-Madison), the Purdue Research Foundation, and the Xerox Corporation. This paper was typeset at Xerox PARC using TEX in the Cedar environment.

...-1717/88/010213 + 24 $03.00/0

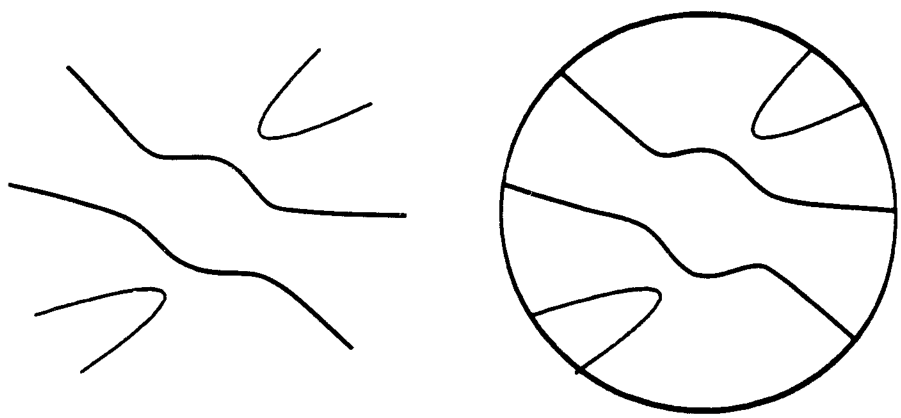


Figure 1: Sample algebraic curve.

to be nonsingular, so the facts cited in the Abstract tell us that C has at most four connected components, each of which is an oval. Recalling that antipodal boundary points are identified in the disk model of the projective plane, we can guess from Fig. 1 that C has two components, but it may not be obvious whether one includes the other or not. Given this particular $f(x, y, z)$ as input, the algorithm we present in this paper determines that C has two ovals, one included in the other.

Our algorithm divides naturally into two main steps. To describe them we use the notion of a cellular decomposition (cd) of a topological space. Let us recall the limited form of it we need here (cf. Massey, 1978, p. 54ff.). For any $i \geq 0$, an *i-cell* is essentially (to be precise, is homeomorphic to) an i-dimensional open ball. Thus a 0-cell is a point, a 1-cell is an "open arc", a 2-cell an "open region", etc. Let X be a subset of the projective plane RP^2; we can view X as a topological space with the topology it inherits from RP^2. A *cellular decomposition* of X is a nested sequence $X^0 \subset X^1 \subset X^2 = X$ of closed subspaces, such that X^0 consists of finitely many 0-cells, $X^1 - X^0$ consists of finitely many disjoint 1-cells, and $X^2 - X^1$ consists of finitely many disjoint 2-cells. Given a cd D of X, and a subset Y of X, we say that D is *compatible with* Y if Y is the union of certain cells of D. Fig. 2 shows a cd of the projective plane compatible with the curve that we looked at in Fig. 1. This cd consists of eleven 0-cells, twenty-three 1-cells, and thirteen 2-cells.

We will be much concerned with the precise arrangement of cells in the cd's we work with. Informally, two (distinct) cells of a cd are *adjacent* if they touch; formally, this is the condition that their union be connected. Clearly adjacency is a symmetric relation on a cd. Thus, we can represent it as an undirected graph which has a vertex for each cell of the cd, and an edge between every pair of adjacent cells. This is the *connectivity graph* of the cd, and is the basic data structure for our algorithm.

The first main step of our algorithm, the subject of Sections 3-6, starts with the input polynomial $f(x, y, z)$, determines whether the curve C defined by f is nonsingular, and if travel along an affine line towards infinity, the extended lines we have defined are ...logical circles.

...n same spirit, what is considered to be a true algebraic curve will be obtained by ...ng certain of the points at infinity to an affine curve. First let us join the affine ...e and the points at infinity into a new object: the real projective plane RP^2 is the ...affine plane extended by the collection of all points at infinity. It is standard to ...del RP^2 as a closed disk (see e.g. Kendig, 1978, p. 6), with antipodal points on the ...ndary identified. The open interior of the disk corresponds to the affine plane, and ...boundary circle to the points at infinity. Even though we are thinking of antipodal ...s of points as being the same point of the projective plane, the boundary circle is ...topologically a circle in RP^2. It is customarily referred to as the *line at infinity* and ...oted l_∞. Naturally any particular point of l_∞ is the point at infinity for the affine ...through the origin which approaches it, and for all affine lines parallel to that line. ...Since RP^2 has more points than the affine plane, to give its points coordinates we ...t somehow expand the coordinate system we used for the affine plane. We do so ...h homogeneous coordinates. We start with the convention that each point (x, y) in ...affine plane is assigned the coordinate triple $[x, y, 1]$ as a point of the projective ...ne. We further adopt the convention that for any nonzero real number t, all triples ...$ty, t]$ correspond to exactly the same point of the projective plane (so a point in the ...jective plane has a whole equivalence class of coordiate triples, any one of which is ..."homogeneous coordinates"). By definition, there is no point in the projective plane ...h coordinate triple $[0, 0, 0]$. As we saw, any point at infinity is the limit point of ...e affine line through the origin. Consider an affine line through the origin which ...contains the affine point (a, b) different from the origin. All points on the line are ...he form (ta, tb), for t real. Hence, as we travel out to infinity along the line, the ...jective coordinates of the points we pass over can be written in the form $[a, b, 1/t]$ ...t steadily decreasing (in absolute value). Hence it is natural to assign the coordinate ...ple $[a, b, 0]$ to the point at infinity by which we extend this line. Note that with this ...ordinate assignment, we indeed approach the same point at infinity whether we travel ...t to infinity along the line by making t large positive or large negative.

In order to have algebraic curves in the projective plane, we must have polynomial ...uations which define them. Since points in the projective plane have coordinate triples ...$y, z]$, we need trivariate polynomials $f(x, y, z)$. Furthermore, it must be the case that ...$x, y, z) = 0$ if and only if $f(tx, ty, tz) = 0$ for nonzero t. We achieve this by limiting ...urselves to *homogeneous* polynomials, i.e. polynomials in which every monomial has ...e same total degree. For example, the polynomial $y^2 - x^2 + xz - z^2$ is homogeneous, ...t $y^2 - x^2 + x - 1$ is not. It is easy to see that any homogeneous f has the property ...hat $f(x, y, z) = 0$ if and only if $f(tx, ty, tz) = 0$ for nonzero t.

Now suppose we have a defining polynomial $g(x, y)$ for an affine curve, say $y^2 - x^2 + x - 1$, and suppose we construct the "homogenization" of it, namely $f(x, y, z) = y^2 - x^2 + xz - z^2$. Clearly, for any point (a, b) of the affine plane, $g(a, b) = 0$ if and only if $f(a, b, 1) = 0$. Hence, in the affine plane, f defines the same curve as g. As one might expect, the manner in which we want to extend affine curves is so that they contain their limit points on l_∞. It turns out that all such limit points will be solutions of the equation $f(x, y, z) = 0$ which lie on l_∞. As we now show, this follows from reasoning which is essentially a repetition of our original derivation of homogenous coordinates. If we are approaching infinity on an affine curve, then in projective coordinates, we have points $[x, y, 1]$ in which either x, or y, or both, are approaching infinity. Suppose without

loss of generality that y is. We can represent the points we are looking at as [x/
Hence as y approaches infinity, we approach the point $[x/y, 1, 0]$. By a straig
continuity argument, the fact that all points $[x/y, 1, 1/y]$ satisfy $f = 0$, im
$f(x/y, 1, 0) = 0$. Thus every point we want to add to our affine curve is inde
the solutions of $f = 0$.

We define an algebraic curve in the projective plane (a *projective cu*
the point set (in the projective plane) defined by a homogeneous polynomial
$f(x, y, z) = 0$. Obviously for any homogeneous $f(x, y, z)$, we can get a certain
$g(x, y)$ (which in general is not homogeneous) by evaluating f at $z = 1$. g de
tain affine curve, which we call an "affine representative" of the projective cur
by f. We say that the curve $f = 0$ is the "projective completion" of the af
defined by g. The projective completions of affine lines are just the extended
we began with above.

Let C be the projective curve defined by $f(x, y, z) = 0$ for some f. A point o
jective plane, with homogeneous coordinates $[x, y, z]$, is a ***singular*** point of C i
$= f_x(x, y, z) = f_y(x, y, z) = f_z(x, y, z) = 0$ (f_w denotes the partial derivative
respect to w). C is *nonsingular* if it has no singular points. Walker (1951, p
pictures of some of the different kinds of singularities which can arise. Project
can have points on l_∞ which are not limit points of the curve's affine repre
these are always (isolated) singular points of the curve. Where $g(x, y) = f(x,$
singular points of f in the affine plane correspond to the common zeros of g,
(see Walker, 1951, p. 54).

Pictures of singular curves such as Walker's may serve to make plausible
that a nonsingular projective curve is a compact one dimensional manifold
1978). This captures such observations as "nonsingular curves do not cross the
and "nonsingular curves do not have isolated points", that the pictures su
compact one-dimensional manifold is known to be homeomorphic to a disjoint
topological circles (Milnor, 1965). A topological circle is known to have two
imbeddings in the projective plane (Wilson, 1978). One is what we called an o
other is like the imbedding of a projective line in the projective plane (e.g. l_∞
second kind of imbedding in the projective plane). Harnack's theorem (Wilson
which we cited in the Abstract, established that a projective curve of degree
most $\frac{(n-1)(n-2)}{2} + 1$ connected components. It is known that for a nonsingular
even degree, each component is an oval (Wilson, 1978). For a nonsingular curv
degree, one component is like a projective line, and all the rest are ovals.

It is illustrative to consider nonsingular curves of degree two, i.e. conics. Re
in the affine plane we have a variety of nonsingular conics, i.e. parabolas, hype
circles, ellipses. The results we have cited say that a nonsingular projective co
at most $\frac{(2-1)(2-2)}{2} + 1 = 1$ components, and since degree two is even, this compo
an oval. Thus the projective completion of any nonsingular affine conic is an ov
projective plane. Consider, for example, the hyperbola $xy - 1 = 0$, which in th
plane has two "branches": its projective completion is the nonsingular projectiv
$xy - z = 0$. Setting $z = 0$, we see that it has the points $[1, 0, 0]$ and $[0, 1, 0]$ on l_∞
the two branches "connect up" on l_∞, and so in the projective plane the curve c
of a single oval.

Projective curves C_1 and C_2 have the same *topological type* if there is a hom
phism of the projective plane to itself which maps C_1 onto C_2. If C is a nons
curve defined by $f(x, y, z) = 0$, then the parity of the degree of f, the number o

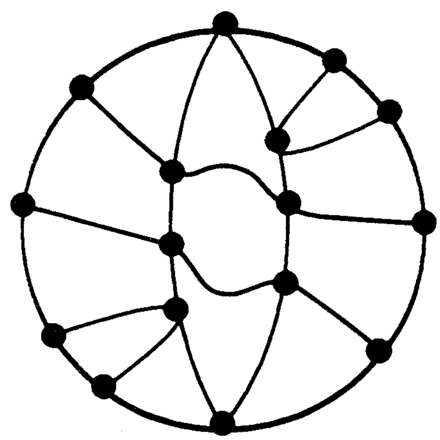


Figure 2: Cellular decomposition of projective plane compatible with sample algebraic curve.

so, constructs a (connectivity graph for a) cd of the projective plane. The key component here is construction of a certain *cylindrical algebraic decomposition* (cad) of the affine plane $z = 1$ (see e.g. Arnon *et al.*, 1984a, for information on cad's). This cad is a cd of the affine plane, and the cad algorithm from Arnon *et al.* (1984a, 1984b) that we use constructs its connectivity graph, so the only work that remains for us is to extend it to a cd of the projective plane and construct the enlarged connectivity graph. Fig. 2 shows the cd of the projective plane our algorithm produces for the sample $f(x, y, z)$ given at the beginning of this Introduction. Identification of the cells of this cd that belong to C is straightforward.

The second of the two main steps of our algorithm, discussed in Section 7, assumes that the input polynomial $f(x, y, z)$ defines a nonsingular curve C, that a cellular decomposition D^* of the projective plane compatible with C has been constructed, that D^*'s connectivity graph has been constructed, and that the cells (vertices) in the connectivity graph which belong to C are marked in some fashion. The topological type, i.e. the number and ordering of ovals, of C is then computed. The key idea is to reduce the determination of the ovals and their ordering to a series of connected components and Euler characteristic computations in appropriate subgraphs of the connectivity graph.

Section 8 summarizes our discussion with a main algorithm TOPTYP, and in Section 9 we trace TOPTYP for the sample $f(x, y, z)$ that we considered above. Section 10 contains an analysis of the TOPTYP's maximum computing time, which we show to be $O(n^{27}L(d)^3)$. This is not out of character with other algebraic algorithms, such as polynomial factorization (Kaltofen, 1982), and indeed the most costly parts of our algorithm are standard algebraic algorithms for such tasks as algebraic number computations, root isolation, and resultant computation. And as for other algebraic algorithms, despite a high worst-case computing time bound, our topological type algorithm has a useful range of application in practice. It has been implemented and applied to examples such as the one considered above.

We summarize basic properties of algebraic curves and homogeneous coordinates, and provide background on the facts cited in the Abstract, in Section 2. The purpose of this material is to establish connections between the conventions and viewpoints of algebraic geometry and those of computer algebra. The knowledgeable reader may skim or skip it.

We were led to seek a topological type algorithm by remarks of Gudkov (1974, p. 67). Polotovskii (1973) gave a topological type algorithm for so-called 'rough' curves of even degree, but did not discuss its feasibility or computing time. His approach is quite different from ours: he examines the curves $f(x, y, z) + \epsilon z^n$, ($n = degree(f)$), for various small values of ϵ. We have recently learned of an independently developed topological type algorithm by P. Gianni and C. Traverso (1983), which has some resemblance to our method, but does not make use of cellular decompositions. As noted by Fuks (1974), one could get a topological type algorithm directly from a decision procedure for elementary algebra and geometry (e.g. Tarski, 1951, Collins, 1975, or Ben-Or *et al.*, 1986), but it seems unlikely that such an algorithm would have a polynomial time bound.

The algorithm we give in this paper existed in rough form by summer 1982, and was presented in a seminar at Purdue University in February 1983. An expanded version of this paper has appeared as Arnon & McCallum (1983), and an abstract of it as Arnon & McCallum (1984). A graph data structure for cad's similar to the one we use in this paper is employed in the cluster-based cad algorithm (Arnon, 1988).

2 Algebraic curves

Kendig (1978) and Walker (1951) provide additional coverage of the material contained in this section. ***Real affine space*** is euclidean space without the notions of distance and angle (Walker, 1951). An important difference between euclidean and affine space is that the usual distance function gives euclidean space a (metric space) topology, which affine space lacks. The cylindrical algebraic decomposition algorithm makes essential use of this topological structure of euclidean space. In this paper, the term "affine space" is mainly useful to us as a means of referring to a certain distinguished subset of projective space. We view affine and euclidean space as essentially the same, i.e. we take for granted the existence of the usual topology on real affine space whenever we need it. We write E^i to denote i-dimensional euclidean space.

Curves in the affine plane defined by polynomial equations $g(x, y) = 0$ are called ***affine algebraic curves***. Throughout this paper, when we say "curve", we mean what is usually referred to as a "real curve", i.e. the real (and not, for example, the complex) solutions to a polynomial equation such as $g(x, y) = 0$. It has long been the prevailing viewpoint in algebraic geometry that an affine curve is only a portion of some true algebraic curve. This point of view is motivated by consideration of the intersections of lines in the affine plane. Two lines in the affine plane have exactly one intersection, unless they are parallel, in which case they do not intersect. We can remove this exceptional behavior of parallel lines by extending affine lines to contain a "point at infinity", and specifying that lines which are parallel in the affine plane have the same point at infinity. Now we can say that any two (extended) lines, without exception, have exactly one intersection. Since we get to the same point at infinity no matter which direction has, and the partial ordering of its ovals induced by the inclusion relation, com- determine its topological type (this result depends upon elementary facts from mensional topology, for example, the Schoenflies theorem: see Moise, 1977). Thus orithm we give in this paper produces information which completely determines ological type of a nonsingular curve.

1900, Hilbert posed the problem (known since as Hilbert's Sixteenth Problem) of ining which logically possible topological types of a nonsingular curve of degree n lised by an actual curve. It is known that the number which actually occur are a raction of the number which are logically possible, but the problem is still largely ed. Curves with the maximum possible number of ovals (so-called M-curves), of egree, are of greatest interest. Obviously there is only one type of M-curve of two. It is easy to show that the only type of M-curve of degree four is four ovals rnal to each other. Note that the sample $f(x, y, z)$ we looked at in Section 1 has four but only two ovals, and so is not an M-curve. Hilbert knew that there could more than three types of M-curves of degree six; the proof that there are exactly as not completed until 1967. For degree eight and up the problem is unsolved. lson (1978) and Gudkov (1974) for further information on Hilbert's Sixteenth m.

Preliminaries

king up the essential ideas of the first main step of our algorithm we attend to liminaries. As before, C denotes the curve defined by $f = 0$, where $f(x, y, z)$ is polynomial to our algorithm. First, we will assume that $f(x, y, z)$ is squarefree no multiple factors). If it is not, we may replace it by its greatest squarefree (x, y, z) (h is f divided by $gcd(f, f_z)$), since the curve defined by $h = 0$ is just C. ofen (1982, p. 98), or Collins & Loos, (1982, p. 84), for squarefree factorization ms. Second, assuming the convention that the affine plane is the set of points of the form $[x, y, 1]$ and that l_∞ is all points of the form $[x, y, 0]$, we want f to e following conditions:

C has only finitely many points on l_∞, each of which is simple (this is equivalent quirement that $f(x, y, 0)$ be nonzero and squarefree);

The point $[0, 1, 0]$ of l_∞ does not lie on C.

rpose of these conditions is to simplify determination of adjacencies between e affine plane and cells on l_∞, the subject of Section 6. Note also that if C (C1) and (C2), then it has no singularities on l_∞. We now show that if (C1) are not satisfied initially, then we can perform a linear change of coordinates of h transforms $f(x, y, z)$ to a polynomial $F(U, V, W)$, such that F is squarefree ogeneous of the same degree as f, C is nonsingular if and only if the curve d by $F = 0$ is nonsingular, C and C' have the same topological type, and C' onditions (C1) and (C2).

rategy is to find a line that has only simple intersections with C, and to change es so that that line becomes the line at infinity. We need to work with a num- ferent homogeneous polynomials $p(x, y, z)$ in our derivation of this coordinate Thus, in this section only, we will write C_p to denote the real projective curve defined by $p(x, y, z)$. We may assume that the input polynomial $f(x, y, z)$ is squarefree. Let n be the degree of f, and let

$$f(x, y, z) = f_r(x, y,)z^{n-r} + \dots + f_n(x, y),$$

where $0 \leq r \leq n$, each $f_i(x, y)$ is homogeneous of degree i, and $f_r(x, y) \neq 0$. Suppose $f_n(x, y) = 0$. Then z divides $f(x, y, z)$, but z^2 doesn't divide $f(x, y, z)$, since $f(x, y, z)$ is squarefree. We can therefore write $f(x, y, z) = z\ \phi(x, y, z)$, where

$$\phi(x, y, z) = f_r(x, y)z^{n-r-1} + \dots + f_{n-1}(x, y)$$

and $f_{n-1}(x, y) \neq 0$. l_∞ is contained in the curve C_f, hence if C_ϕ has any point on l_∞ (that is, if either $f_{n-1}(0, 1) = 0$, or $f_{n-1}(1, y)$ has a real root), then C_f is singular, we report this fact, and exit from the algorithm. If C_ϕ does not meet l_∞, then C_f is nonsingular if and only if C_ϕ is nonsingular. Moreover, if C_ϕ is nonsingular, then C_f and C_ϕ have the same number and arrangement of ovals. Hence we can replace f by ϕ; since C_ϕ does not meet l_∞, conditions (C1) and (C2) are trivially satisfied.

Suppose now that $f_n(x, y) \neq 0$. Let us transform $f(x, y, z)$ to $F(X, Y, Z)$, such that $F(0, 1, 0) \neq 0$ (so that the point [0,1,0] is not on C_F). We know $f_n(x, 1) \neq 0$, since otherwise $f_n(x, y) = 0$. Thus there is an integer λ such that $f_n(\lambda, 1) \neq 0$. Define $F(X, Y, Z)$ by

$$F(X, Y, Z) = f(X + \lambda Y, Y, Z)$$

Then $F(0, 1, 0) = f(\lambda, 1, 0) = f_n(\lambda, 1) \neq 0$. Let $G(X, Y) = F(X, Y, 1) \neq 0$ and let $D(X)$ be the discriminant of $G(X, Y)$. Then $D(X) \neq 0$, since $G(X, Y)$ is squarefree (and nonzero). Find an integer κ with $D(\kappa) \neq 0$. Change variables as follows: $X = W + \kappa U$, $Y = V$, $Z = U$. Since $W = X - \kappa Z$, the line $X = \kappa Z$ (which is the projective version of the affine line $X = \kappa$) corresponds to the line $W = 0$ (which is the line at infinity in U, V, W coordinates). Let

$$E(U, V, W) = F(W + \kappa U,\ V,\ U).$$

E is squarefree and homogeneous of the same degree as f. Observe that $E(0, 1, 0) = F(0, 1, 0) \neq 0$. We have $E(U, V, 0) = F(\kappa U, V, U)$, so that $E(1, V, 0) = F(\kappa, V, 1) = G(\kappa, V)$, a nonzero squarefree polynomial (since $D(\kappa) \neq 0$). Thus $E(U, V, 0)$ is nonzero and squarefree. Hence C_E satisfies conditions (C1) and (C2).

It remains to show that (i) C_f is nonsingular if and only if C_E is nonsingular; and (ii) C_f and C_E have the same topological type. Let $T(x, y, z) = (z, y, x - \kappa z - \lambda y)$. Then T is an invertible linear transformation of E^3 with inverse $T^{-1}(U, V, W) = (W + \kappa U + \lambda V, V, U)$. We have

$$E(U, V, W) = f(T^{-1}(U, V, W)) .$$

Applying the chain rule for differentiation, we obtain

$$\begin{pmatrix} E_U \\ E_V \\ E_W \end{pmatrix} = \begin{pmatrix} \kappa & 0 & 1 \\ \lambda & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix}$$

The matrix on the right side of this equation is invertible. Hence by the two preceding equations, (U, V, W) is a singular point of C_E if and only if $T^{-1}(U, V, W)$ is a singular

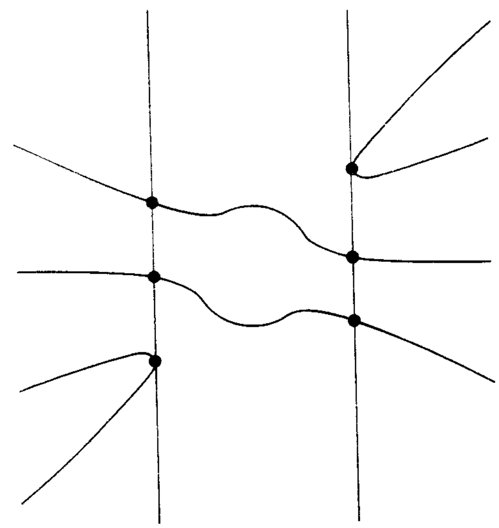


Figure 3: Sample cylindrical algebraic decomposition of E^2.

point of C_f. This proves (i). Since T is an invertible linear transformation of E^3, T induces a homeomorphism $\tilde{T}: RP^2 \to RP^2$ given by $\tilde{T}[x, y, z] = [T(x, y, z)]$. Clearly $\tilde{T}$ carries C_f onto C_E. Thus C_f and C_E have the same topological type: so (ii) is proved.

The reader may wonder why we do not transform C_f to a curve which has no intersections with l_∞. Ragsdale (1906, p. 377 footnote) notes that there exist curves which, for any linear change of coordinates, will have points on l_∞.

4 Decomposition and adjacencies of the affine plane

We will discuss testing the curve C for the presence of singularities later in this section. Assume for the moment that we have determined that C is nonsingular, and we wish to construct an appropriate cd of the projective plane. As detailed in Section 2, we view the projective plane RP^2 as the disjoint union of an affine plane and a line at infinity. Our strategy is to separately construct decompositions and connectivity graphs for the affine plane and l_∞, and then to join the two connectivity graphs by adding edges that join a cell of one decomposition that is adjacent to a cell of the other. Thus we arrive at a connectivity graph for a decomposition of the projective plane. Both the decomposition of the affine plane and the decomposition of the line at infinity are constructed so as to be compatible with C, and so that the cells belonging to C are marked in some fashion. Thus, on completion of these steps we have a decomposition of the projective plane that is compatible with C, and in which the cells belonging to C are marked. In this section we discuss the decomposition of the affine plane. In Section 5 we discuss the decomposition of the line at infinity, and in Section 6, determination of adjacencies between cells in the affine plane and cells in l_∞.

Assume now that $f(x, y, z)$ is squarefree, homogeneous, and satisfies conditions (C1) and (C2) of Section 3. Let $g(x, y) = f(x, y, 1)$. As mentioned in Section 1, the decomposition we construct of the affine plane is a cylindrical algebraic decomposition (cad). Given input $g(x, y)$, we construct a cad that is ***g-invariant***, i.e. for each cell c of the cad, either $g(x, y) < 0$ for all (x, y) in c, or $g(x, y) = 0$ for all (x, y) in c, or $g(x, y) > 0$ for all (x, y) in c. Fig. 3 shows the cad of E^2 that we construct for the example of Section 1, where $g(x, y) = y^4 - 2xy^3 - x^2y^2 + y^2 + 2x^3y + x^2 - 1$.

Algorithm *AffinePlaneDecomp* given in Fig. 4 presents our complete algorithm for decomposition and adjacencies of the affine plane. It basically consists of portions of algorithm CADA2 of Arnon *et. al.* (1984b). It detects singularities of C in the affine plane as it builds the cad and halts if any are found. The method used for singularity detection is the same used for determination of basis polynomial signatures in the cluster-based cad algorithm (Arnon, 1988). As we have said, the cad that is constructed is a cd; *AffinePlaneDecomp* produces the connectivity graph of this cd and marks the vertices (cells) in it that belong to C. The map *gsfd* used in the algorithm denotes "greatest squarefree divisor", and the map *PROJ* used in the algorithm is as defined in Arnon *et. al.* (1984a).

5 Decomposition and adjacencies of the line at infinity

We isolate the roots of $f(1, y, 0)$, a univariate polynomial with integer coefficients (Collins & Loos, 1982). For each root y, $[1, y, 0]$ is a point of C on l_∞, and we make it a 0-cell of our decomposition of l_∞. We make also the point $[0, 1, 0]$ (which is not on C by condition (C2) of Section 3) a 0-cell of the decomposition. We then take the complementary open intervals of these 0-cells to be the 1-cells of our decomposition of l_∞. The cells on the boundary circle of the disk in Fig. 2 illustrate these steps for the sample curve of Section 1. Recalling the identification of antipodal points in the disk model of RP^2, we see that l_∞ is decomposed into five 0-cells and five 1-cells in this example. In sum, once we know how many roots $f(1, y, 0)$ has, we know what the cells of the decomposition of l_∞ are, and what the connectivity graph of this decomposition is.

Let us introduce some notation for the cells of our decomposition of l_∞. Suppose that there are $k \geq 0$ points of C on l_∞. Since $[0, 1, 0]$ is not on C, these points can be written $[1, \gamma_1, 0] = P_1, \ldots, [1, \gamma_k, 0] = P_k$, where $\gamma_1 < \ldots < \gamma_k$ are the real roots of $f(1, y, 0)$. Our cellular decomposition of l_∞ consists of: $P_1, \ldots, P_k$, the point $[0, 1, 0] = P_0 = P_{k+1}$, and the 1-cells e_i, $0 \leq i \leq k$, where e_i is the open interval in l_∞ bounded by P_i and P_{i+1}. Fig. 5 illustrates this notation for the sample curve of Section 1. Let us assign to P_i the index $(0, 2i)$, and to e_i the index $(0, 2i + 1)$, for $0 \leq i \leq k$. This assignment of indices preserves the rule that the dimension of a cell is the sum of the parities of the components of its index (cell indices are defined in Section 6).

$G \leftarrow$ **AffinePlaneDecomp** ($g(x,y)$)

Input: $g(x,y)$ is a primitive bivariate polynomial with integer coefficients. Let C_g denote the curve in the real affine plane defined by $g = 0$.

Outputs: If C_g is nonsingular, then G is the connectivity graph for a g-invariant cad D of the Euclidean plane in which the cells comprising C_g are marked. If C_g is singular, then T is the string "SINGULAR".

(1) [Base case.] Set $P \leftarrow PROJ(\{g\})$. Isolate the real roots of the irreducible factors of the nonzero elements of P to determine a cad D' of the real affine line. Construct a sample point for each cell of D' as in algorithm CADA2 of Arnon *et. al.* (1984b).

(2) [Extension and singularity check.] Let $a_1 < a_2 < \cdots < a_{2m} < a_{2m+1}$, $m \geq 0$, be the sample points for D' (each a_{2i+1} is a rational sample point for a 1-cell; each a_{2i} is an algebraic sample point for a 0-cell). For $i = 1, \ldots, 2m+1$, let c_i denote the cell of D' whose sample point is a_i, and do the following five things: first, construct open isolating intervals for the real roots of $g(a_i, y)$ to determine the sections of a stack $S(c_i)$ in E^2; second, compare the signs of $gsfd(g_x(a_i, y))$ at the endpoints of each isolating interval, and the signs of $gsfd(g_y(a_i, y))$ at the endpoints of each isolating interval, and if the signs are different in either comparison, then set $G \leftarrow$ "SINGULAR" and exit; third, construct cell indices for the cells of $S(c_i)$ and add a vertex for each cell of $S(c_i)$ to G; fourth, mark each section of $S(c_i)$ as belonging to C_g; fifth, add the intrastack adjacencies for $S(c_i)$ to G. When we are done, we have built a g-invariant cad D of the affine plane that is *proper* in the sense of Arnon *et. al.* (1984b).

(3) [Adjacency computation.] For $i = 1, \ldots, m$, call algorithm SSADJ2 of Arnon *et. al.* (1984b) with inputs g, a_{2i}, a_{2i-1}, and a_{2i+1}, and add the contents of its outputs L_1 and L_2 to G. Note that the section numbers which occur in the adjacencies returned by SSADJ2 must first be converted into the indices of the corresponding cells of D; for example, if the list L_1 returned by the i^{th} call to SSADJ2 contains the adjacency $\{3, 2\}$, it must be converted to $\{(2i, 6), (2i-1, 4)\}$ before being added to G. Infer the remaining interstack adjacencies between $S(c_{2i})$ and $S(c_{2i-1})$, and between $S(c_{2i})$ and $S(c_{2i+1})$, as described at the end of Section 2 of Arnon *et al.* (1984b), and add them to G. All adjacencies of D are now recorded in G. □

Figure 4: Algorithm AffinePlaneDecomp.

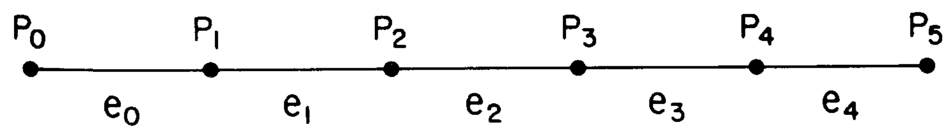


Figure 5: Cellular decomposition of l_∞ for sample curve.

6 Adjacencies between finite and infinite cells

Let us now consider the question of determining the adjacencies between a cell in our decomposition of the affine plane (which we call a *finite* cell) and a cell in our decomposition of the line at infinity (an *infinite* cell). Consider the sample curve C shown in Figs. 1 and 2. We see that for this example it is straightforward to describe all such adjacencies. The affine 1-cells which are on C, and which "go off to infinity" to the "right" or to the "left", are each adjacent to exactly one 0-cell on the line at infinity, and this 0-cell is not $[0,1,0]$, In fact, for this example, when we have either determined how many roots $f(x,1,0)$ has, or found out how many finite 1-cells that lie in C go off to infinity to the "right", or found out how many finite 1-cells that lie in C go off to infinity to the "left", we know exactly what the adjacencies among finite 1-cells that lie in C and infinite 0-cells are. Furthermore, all cells which are topmost or bottommost in some stack of the cad, and only those cells of the cad, are adjacent to the point [0,1,0] on the line at infinity. This is precisely the situation that conditions (C1) and (C2) of Section 3 are designed to bring about. We prove with Theorem 6.1 below that they do.

We first review some definitions from Arnon *et al.* (1984a). Let X be a nonempty connected subset of E^1. The *cylinder over* X, written $Z(X)$, is $X \times E$. A *section* of $Z(X)$ is a set s of points $< x, \phi(x) >$, where x ranges over X, and ϕ is a continuous, real-valued function on X. s, in other words, is the graph of ϕ. The constant functions $\phi = -\infty$ and $\phi = +\infty$ are allowed; in these cases, s is an *infinite section*. A *sector* of $Z(X)$ is a set $\hat{s}$ of all points $< x,\ y >$, where x ranges over X, and $\phi(x) < y < \psi(x)$ for (continuous, real-valued) functions $\phi < \psi$. A *stack* over X is a collection of disjoint sections and sectors of $Z(X)$ whose union is $Z(X)$; X is the *base* of the stack. A *cylindrical decomposition* D of E^2 is a finite cellular decomposition of E^2, such that for some cellular decomposition D' of the real line E^1, the cells of D comprise stacks over the cells of D'. The decomposition shown in Fig. 3 is a cylindrical decomposition D of E^2. For that example, D' consists of two 0-cells and the three complementary 1-cells.

Any cell c of a cylindrical decomposition D of E^2 can be assigned an index, consisting of an ordered pair of positive integers. The first component specifies the cell of D' which is the base of the stack containing c; the cells of D' are numbered 1, 2,... from left to right. The second component specifies the position of c within the stack; the cells of the stack are numbered 1, 2,... from bottom to top. The indices for the cells in Fig. 3 are shown in Fig. 6. It is easily seen that the dimension of a cell in a cylindrical decomposition is the sum of the parities (even = 0, odd = 1) of the components of its index, e.g. (1,9) is a 2-cell, (2,6) is a 0-cell.

In a cylindrical decomposition D of E^2, the cells over the leftmost cell of D' comprise the *leftmost stack* of D, and the cells over the rightmost cell of D' the *rightmost stack*. Thus in Fig. 6, the cells with indices $(1,\ j)$, $1 \leq j \leq 9$, are the leftmost stack, and the cells with indices $(5,\ j)$, $1 \leq j \leq 9$, are the rightmost stack. In general, the leftmost and rightmost stacks of D are distinct, however if D' consists of a single 1-cell (the entire real line), then the leftmost and rightmost stacks are identical. This concludes our review of definitions.

THEOREM **6.1** *Let $f(x,y,z)$ be a homogeneous trivariate polynomial with integer coefficients, which satisfies conditions (C1) and (C2) of Section 3. Suppose that $f(1,y,0)$ has $k \geq 0$ real roots $\gamma_1 < \ldots < \gamma_k$. Let $g(x,y) = f(x,y,1)$, and suppose that S and T are (respectively) the rightmost and leftmost stacks of a g-invariant cad of the affine plane. Then*

(1,9)				(5,9)
(1,8)				(5,8)
(1,7)	(2,7)		(4,7)	(5,7)
(1,6)	(2,6		(4,6)	(5,6)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)
(1,1)	(2,1)	(3,1)	(4,1)	(5,1)

Figure 6: Cell indices for sample cylindrical decomposition.

(i) *S has k sections, say $s_1 < \ldots < s_k$, and T has k sections, say $t_1 < \ldots < t_k$;*
(ii) *for $1 \leq i \leq k$, if s_i is the graph of the continuous real-valued function ϕ_i, and t_i the graph of the continuous real-valued function ω_i, then*

$$\lim_{x \to +\infty} \frac{\phi_i(x)}{x} = \gamma_i ,$$

$$\lim_{x \to -\infty} \frac{\omega_i(x)}{x} = \gamma_{k-i+1} .$$

PROOF. Let n be the degree of $f(x, y, z)$. By condition (C1) of Section 3, each γ_i is a simple root of $f(1, y, 0)$. Let $G(X, Y) = f(1, Y, X)$. Since $f(0, 1, 0) \neq 0$ by condition **(C2)**, $G(X, Y) = g_0 Y^n + g_1(X) Y^{n-1} + \ldots + g_n(X)$, for some constant $g_0 \neq 0$ and polynomials $g_1(X), \ldots, g_n(X)$. Since $G(0, Y) = f(1, Y, 0)$, $G(0, Y)$ has exactly k real roots $\gamma_1 < \ldots < \gamma_k$, each of them simple. Hence by root continuity, there is some $\delta > 0$ such that $|X| < \delta$ implies $G(X, Y)$ has exactly k real roots, each of them simple. The i^{th} of these roots approaches γ_i as $|X| \to 0$. Since $g(x, y) = x^n G(1/x, y/x)$ for nonzero x, $g(x, y)$ has k real roots, each simple, for all sufficiently large positive x. Hence S has k sections. A similar argument shows that T has k sections.

For any x in the interval $(\alpha, +\infty)$ on which ϕ_i is defined, $\phi_i(x)$ is the i^{th} real root of $g(x, y)$. Hence, for positive x greater than α, $\phi_i(x)/x$ is the i^{th} real root of $G(1/x, Y)$. Hence, as $x \to +\infty$, $\phi_i(x)/x \to \gamma_i$. For any x in the interval $(-\infty, \beta)$ in which ω_i is defined, $\omega_i(x)$ is the i^{th} real root of $g(x, y)$. Hence, for negative x less than β, $\omega_i(x)/x$ is the $(k - i + 1)^{st}$ real root of $G(1/x, Y)$. Hence as $x \to -\infty$, $\omega_i(x)/x \to \gamma_{k-i+1}$ □

As we have indicated, Figs. 1-3 illustrate Theorem 6.1. In this example, the stacks S and T of the theorem each have four sections. One sees that the asymptotic slope of s_i, namely γ_i, is equal to the asymptotic slope of t_{k-i+1}.

Let us now give a general procedure for determination of adjacencies between finite and infinite cells. First we relate adjacency of cells to the topological notion of boundary. It is easily shown that two cells are adjacent if and only if one contains a limit point of the other. For a subset X of a topological space, the *boundary* of X, written ∂X, is $\bar{X} - X$, where $\bar{X}$ denotes the closure of X. It is not difficult to show that ∂X is the set of all limit points of X which do not belong to X. Thus two cells are adjacent if and only if one meets the boundary of the other.

Now, for some given f, let S and T be as in Theorem 6.1. Let P_i and e_i be as in Section 5. Consider first adjacencies between sections of S and T and cells in l_∞. Suppose $S \neq T$. We claim that P_i is a limit point of the i^{th} section s_i of S. As in the Theorem, let s_i be the graph of a function ϕ_i. Let $[x_i, \phi(x_i), 1]$ be a sequence of points in s_i, with x_i approaching $+\infty$. Then $\lim[x_i, \phi(x_i), 1] = \lim[1, \phi(x_i)/x_i, 1/x_i]$ $= [1, \gamma_i, 0] = P_i$. It can be shown that P_i is in fact the unique limit point of s_i on l_∞. Similarly, P_{k-i+1} is the unique limit point of t_i on l_∞. If $S = T$, then $s_i = t_i$ has the limit points P_i and P_{k-i+1} in l_∞ (P_i and P_{k-i+1} may coincide).

Consider now sectors of S and T. Suppose the 1-cell c in the induced cad of the line is the base of S, and let $s_0 = c \times \{-\infty\}$ and $s_{k+1} = c \times \{+\infty\}$ denote the infinite sections of $Z(c)$. For $0 \leq i \leq k$, let $\hat{s}_i$ denote the sector of S between s_i and s_{i+1} (similar definitions hold for T by replacing s by t throughout). Note that $\bar{e}_i = e_i \cup \{P_i, P_{i+1}\}$. If $S \neq T$, it is evident that the portion of the boundary of $\hat{s}_i$ contained in l_∞ is $\bar{e}_i$, while the portion of the boundary of $\hat{t}_i$ contained in l_∞ is $\bar{e}_{k-i}$, for $0 \leq i \leq k$ (see example in Fig. 7). If $S = T$ is the only stack of D, the portion of the boundary of $\hat{s}_i$ contained in l_∞ is $\bar{e}_i \cup \bar{e}_{k-i}$ (see example in Fig. 8).

Now let R be any stack of D besides S and T. Let $r_1 < \ldots < r_l$ be the finite sections of R. Let $\hat{r}_0$ and $\hat{r}_l$ be defined as were $\hat{s}_0$ and $\hat{s}_k$. From the disk model for RP^2, it is evident that $\hat{r}_0$ and $\hat{r}_l$ are the only cells of R to have limit points in l_∞, and each in fact does have the unique limit point P_0 in l_∞ (Fig. 2 illustrates this discussion). This completes the determination of adjacencies between finite and infinite cells.

We omit the proof, in the present paper, that the decomposition D^* of the projective plane we construct is actually a cellular decomposition in the sense of Massey (1978). A stronger result can in fact be shown: D^* gives RP^2 the structure of a cell complex. The proof is given in Arnon & McCallum (1983).

7 Topological type from cellular decomposition

In this section, we will think of the cd D^* of RP^2 as a certain collection of cells. Thus we write the connectivity graph as $G^* = (D^*, E)$, where E, the set of edges of G^*, is the set of all pairs of adjacent cells of D^*. We write D_C to denote the subset of D^* consisting of all cells contained in the curve C.

Our first task is to determine the components of C. This is accomplished by constructing the connected components of the subgraph of G^* induced by D_C. In the data structure for G^*, we mark each cell of D_C with an index identifying the particular component of C to which it belongs.

Now, for each component of C, we want to determine if it is an oval, i.e. if its complement with respect to the projective plane has two rather than one components. If so, we want to identify which component of its complement is its interior, and which its exterior. Let $D_1 \subset D_C$ be the cells which comprise a component C_1 of C. We compute the connected components of the subgraph of G^* induced by $D^* - D_1$ (call this subgraph G_1). If there is only one such component, than C_1 is not an oval, and we do no further processing for it.

The Euler characteristic χ, of a cd of a subspace of the projective plane, is $\alpha_0 - \alpha_1 + \alpha_2$, where α_i is the number of i-cells in the cd (cf. Massey, 1978, p. 61). Suppose

Thus two cells are adjacent if and

Theorem 6.1. Let P_i and e_i be as
ections of S and T and cells in l_∞.
int of the i^{th} section s_i of S. As in
Let x_i, $\phi(x_i)$, 1 be a sequence of
$\phi(x_i)$, 1 = lim 1, $\phi(x_i)/x_i$, $1/x_i$
the unique limit point of s_i on l_∞.
l_∞. If $S = T$ then $s_i = t_i$ has the
ay coincide).

cell c in the induced cad of the line
$\{\pm\infty\}$ denote the infinite
r of S between s_i and s_{i+1} (similar
. Note that $\bar{e}_i = e_i \cup \{P_i, P_{i+1}\}$.
ary of $\hat{s}_i$ contained in l_∞ is $\bar{e}_i$, while
e_{k-1}, for $0 \le i \le k$ (see example in
on of the boundary of $\hat{s}_i$ contained

$r_1 < \ldots < r_l$ be the finite sections
rom the disk model for RP^2, it is
ve limit points in l_∞, and each in
2 illustrates this discussion). This
inite and infinite cells.
lecomposition D^* of the projective
n in the sense of Massey (1978).
^{2} the structure of a cell complex.

r decomposition

a certain collection of cells. Thus
ere E, the set of edges of G^*, is
D_C to denote the subset of D^*

. This is accomplished by con-
of G^* induced by D_C. In the
index identifying the particular

mine if it is an oval, i.e. if its
wo rather than one components.
lement is its interior, and which
se a component C_1 of C. We
G^* induced by $D^* - D_1$ (call this
in C is not an oval, and we do

he projective plane, is $\alpha_0 - \alpha_1$
Massey, 1978, p. 61). Suppose

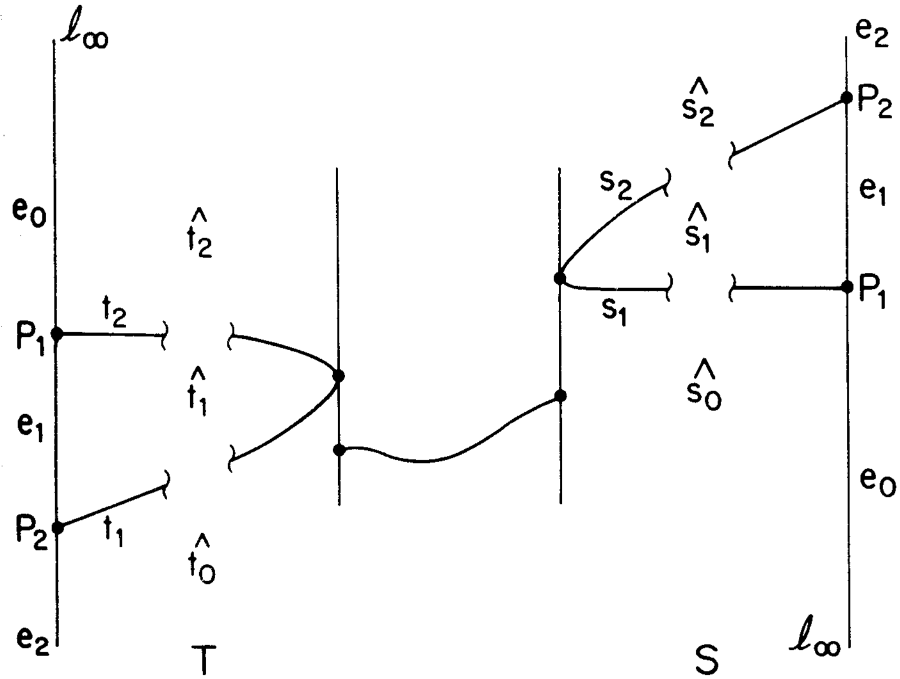


Figure 7: Example of adjacencies between S, T, and l_∞.

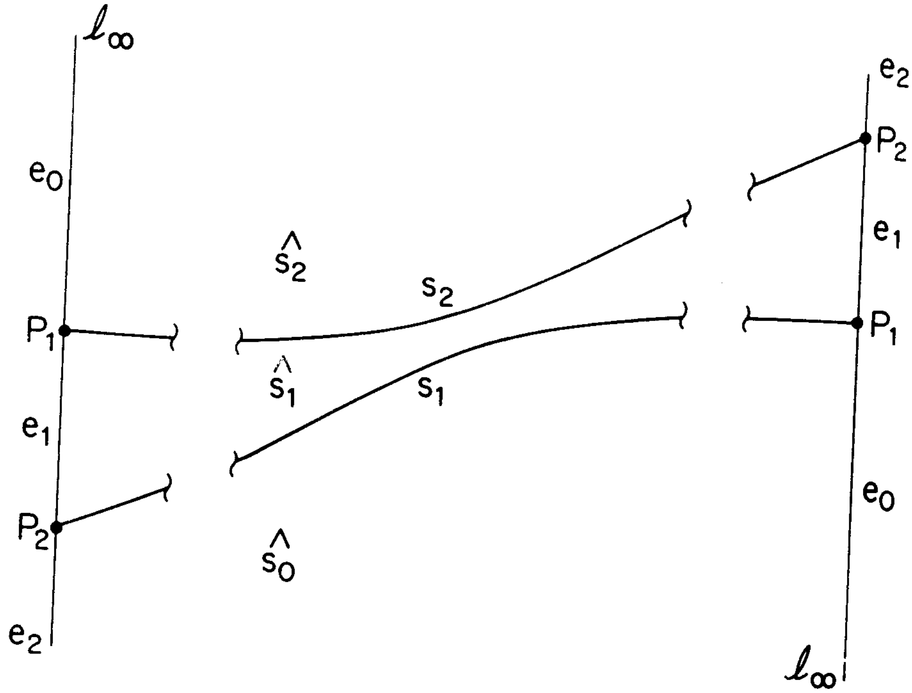
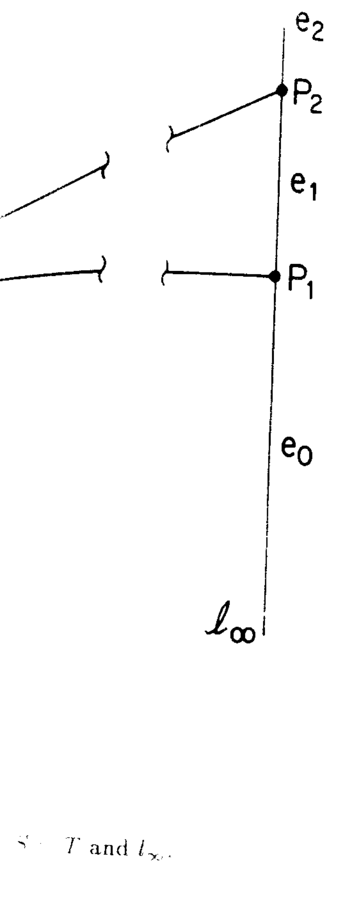


Figure 8: Example of adjacencies between $S = T$ and l_∞.



T and l_∞.

f C_1 is an oval, i.e. G_1 has two connected components. Since D^* is compatible h C, it follows that one of the components of G_1 is a cellular decomposition of the erior of C_1, and the other is a cellular decomposition of the exterior of C_1. As we have ntioned, the interior of C_1 is topologically a disk, and its exterior is topologically a öbius band. It follows (e.g. from Exercise 2 of Massey, 1978, p. 61) that the Euler racteristics of the two components of G_1 must be different. In fact, $\chi(disk) = 1$, d $\chi(\text{Möbius band}) = 0$. Hence, by computing the Euler characteristics of the two mponents of G_1, we determine which is the interior and which the exterior of C_1. ally, in the data structure for G^*, we record at each cell of D^*-D_1 whether it is in e interior or exterior of C_1.

When we have processed each component of C in the above-described fashion, it mains only to determine the partial ordering of C's ovals. We may do this in any mber of ways, for example, by picking one cell in each oval and reading off the order formation we have recorded with it in the data structure for G^*.

Main algorithm

e give a formal description of our main algorithm TOPTYP in Fig. 9. The map *pp* used the algorithm denotes "primitive part" (see Arnon, 1988, for definition of primitive rt).

Example

et us now consider the example of Section 1 in more detail. Let $f(x,y,z)$ be as at the beginning of Section 1. f is irreducible, hence squarefree. $f(x,y,0)$ has no multiple ctors, and [0,1,0] does not lie on C, so we need not change coordinates. Let $g(x,y) = f(x,y,1)$. Recall that the cad of the affine plane $z = 1$ constructed by algorithm CADA2 with input $g(x,y)$ is shown in Fig. 3. We find that C is nonsingular. Continuing, we have

$$f(1,y,0) \;=\; y\,(y-1)\,(y+1)\,(y-2),$$

and so C has the four points [1,0,0], [1,1,0], [1,-1,0], and [1,2,0] on l_∞. Thus the cells of D^* on l_∞ are as shown in Figs. 2 and 5. In Fig. 10 we enlarge Fig. 2 and label each cell with its index. The connectivity graph of D^* is clear from Fig. 10 (or Fig. 2). We find that C has two components, composed respectively of the following collections of cells:

$$J_1 = \{\ (1,2),(2,2),(1,4),(0,6),(5,6),(4,6),(5,8),(0,8)\ \},$$

and

$$J_2 = \{\ (1,6),(2,4),(3,2),(4,2),(5,2),(0,2),(1,8),(2,6),(3,4),(4,4),(5,4),(0,4)\ \}.$$

Since f has even degree, both correspond to ovals.

Let O_1 denote the oval of C that comprises the cells of J_1. *complement*(O_1) turns out to have two components, composed of $K_1 = \{\ (1,3),(0,7),(5,7)\ \}$, and

$$L_1 \;=\; \{\ (0,0),(0,1),(0,2),(0,3),(0,4),(0,5),(0,9),(1,1),(1,5),(1,6),(1,7),$$

$T \leftarrow$ TOPTYP ($f(x,y,z)$)

Input: $f(x,y,z)$ is a homogeneous trivariate polynomial with integer coefficients, i.e. a homogeneous element of $\mathbf{Z}[x,y,z]$. Let C denote the curve in the real projective plane defined by $f = 0$.

Outputs: If C is nonsingular, then T is the number and partial ordering of the ovals of C. If C is singular, then T is the string "SINGULAR".

(1) [Transform f, if necessary.] If f is not squarefree, then replace f by $gsfd(f)$. Test whether $f(x,y,z)$ satisfies conditions (C1) and (C2) of Section 3. If not, then change coordinates as per Section 3 to get some new $f(x,y,z)$. Recall that we may detect during the coordinate change process that C is singular; in this event, set T to the string "SINGULAR" and exit. Set $g(x,y) \leftarrow f(x,y,1)$. By conditions (C1) and (C2), $V(content(g))$ is empty, hence $V(g) = V(pp(g))$, hence replace g by $pp(g)$.

(2) [Decomposition and adjacencies of the affine plane.] Set $G^* \leftarrow AffinePlaneDecomp(g)$. If G^* is the string "SINGULAR" then set T to the string "SINGULAR" and exit.

(3) [Decomposition and adjacencies of the line at infinity.] Determine the number of real roots of $f(1,y,0)$. From this information, place new vertices in the connectivity graph G^* for the corresponding 0- and 1-cells cells on the line at infinity, mark all 0-cells except $[0,1,0]$ as belonging to C, and add new edges to the connectivity graph corresponding to the adjacent pairs of cells within the line at infinity.

(4) [Adjacencies between finite and infinite cells.] As discussed in Section 6, all adjacencies between a 1-cell of the cad that is contained in C and a 0-cell on the line at infinity that is contained in C are now known; add them to the connectivity graph. From these adjacencies, infer the adjacencies of 2-cells of the cad with 0-cells and 1-cells on the line at infinity, and add them to the connectivity graph. Add an edge to the connectivity graph for each adjacency between the point $[0,1,0]$ on the line at infinity, and the topmost and bottommost cell in each stack of the cad of the affine plane. We now have a cellular decomposition D^* of the projective plane that is compatible with C, and all adjacencies of this decomposition are marked in the connnectivity graph.

(5) [Topological type from cellular decomposition.] Determine the components of C, determine which components are ovals, and for each oval, determine which cells of D^* comprise its interior, and which its exterior. From this information, determine the partial ordering of the ovals of C by inclusion. Set T to some representation of the ovals and their partial ordering, and exit □

Figure 9: Algorithm TOPTYP.

y, z) ...

with integer coefficients, i.e. a homo-
n the real projective plane defined by

partial ordering of the ovals of C. If C

en replace f by $gsfd(f)$. Test whether
). If not, then change coordinates as
we may detect during the coordinate
to the string "SINGULAR" and exit.
$(content(g))$ is empty, hence $V(g)$ =

Set $G^* \leftarrow AffinePlaneDecomp(g)$. If
"SINGULAR" and exit.

Determine the number of real roots
in the connectivity graph G^* for the
. mark all 0-cells except $[0,1,0]$ as
graph corresponding to the adjacent

scussed in Section 6, all adjacencies
-cell on the line at infinity that is
ivity graph. From these adjacencies,
d 1-cells on the line at infinity, and
onnectivity graph for each adjacency
opmost and bottommost cell in each
decomposition D^* of the projective
is decomposition are marked in the

ine the components of C, determine
hich cells of D^* comprise its interior.
he partial ordering of the ovals of C
their partial ordering, and exit □

TYP

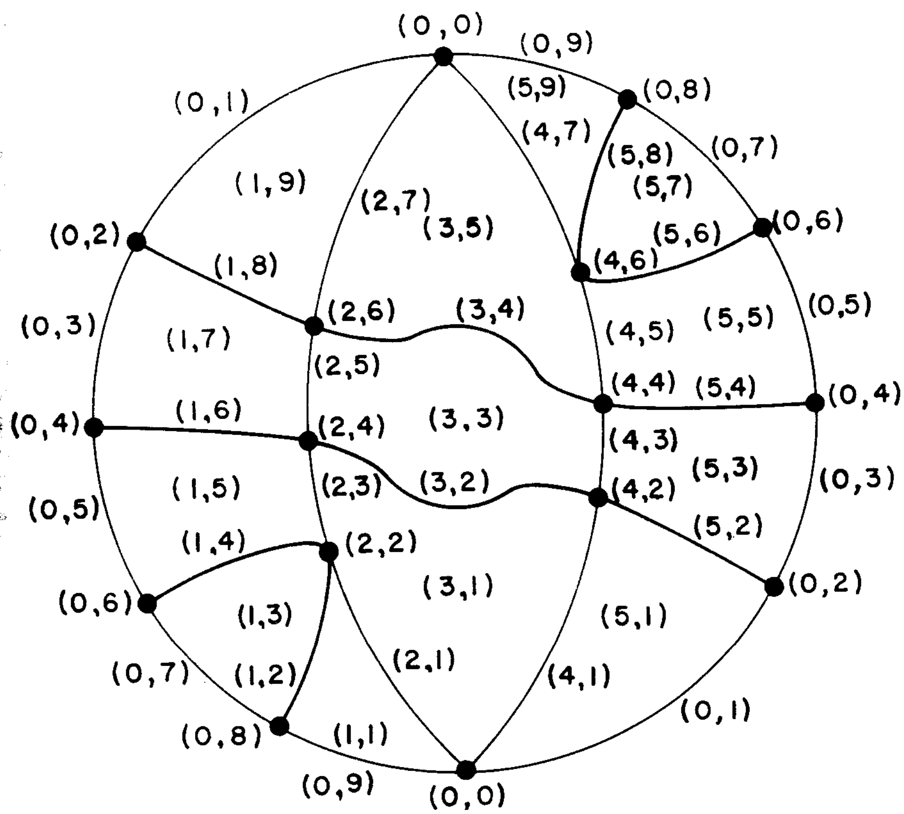


Figure 10: Cellular decomposition of projective plane compatible with sample curve.

$$(1,8),(1,9),(2,1),(2,3),(2,4),(2,5),(2,6),(2,7),(3,1),(3,2),(3,3),(3,4),$$
$$(3,5),(4,1),(4,2),(4,3),(4,4),(4,5),(4,7),(5,1),(5,2),(5,3),(5,4),(5,5),(5,9)\ \}.$$

Computing the dimension of each cell as the sum of the parities of the two components of its index, we find that the Euler characteristic of K_1 is $\chi = 0 - 1 + 2 = 1$. For L_1, we have $\chi = 7 - 18 + 11 = 0$. Hence K_1 corresponds to the interior, and L_1 to the exterior, of O_1.

Consider the second collection J_2 of cells comprising oval O_2 of C. We find two collections of cells corresponding to the components of $complement(O_2)$:

$$K_2 = \{ (1,7),(0,3),(5,3),(4,3),(3,3),(2,5) \}$$

and

$$L_2 = \{ (0,0),(0,1),(0,5),(0,6),(0,7),(0,8),(0,9),(1,1),(1,2),(1,3),(1,4),(1,5),$$
$$(1,9),(2,1),(2,2),(2,3),(2,7),(3,1),(3,5),(4,1),(4,5),(4,6),(4,7),(5,1),(5,5),$$
$$(5,6),(5,7),(5,8),(5,9) \}.$$

The Euler characteristic of K_2 is $\chi = 0 - 3 + 3 = 0$. For L_2, we have $\chi = 5 - 14 + 10 = 1$. Hence K_2 corresponds to the exterior, and L_2 to the interior, of O_2. We see that the cells comprising O_1 occur among the cells comprising the interior of O_2. Hence O_1 is included in O_2, and we have determined the topological type of C.

The time required for this example was approximately five minutes, using an implementation of TOPTYP in the SAC-2 computer algebra system (Collins, 1980) on a Vax 11/785.

10 Computing time analysis

We count the cost of the individual steps of algorithm TOPTYP.

10.1 TOPTYP Step (1)

In computing the greatest squarefree divisor of f, the gcd calculation takes $O(n^5 L(d)^2)$ steps (Collins & Loos, 1982, p. 84; Loos, 1982). This dominates the cost of the division. Using Mignotte's bound (Collins & Loos, 1982, p. 84), the maximum coefficient $\bar{d}$ of the greatest squarefree divisor satisfies $\bar{d} \leq 2^n d$. We do not consider this greatest squarefree divisor calculation to be really a part of our algorithm, and so we will ignore this potential coefficient growth in the remainder of our analysis. Note that the greatest squarefree divisor has lower degree, if different from f.

The significant operations in the coordinate transformation, if it is carried out, are two linear changes of coordinates and computation of the discriminant of a bivariate polynomial of degree n. The cost of the two coordinate changes is dominated by the discriminant computation, and since we will always do a discriminant computation on an input of the same or larger size in step (2), we ignore the cost of the coordinate transformation.

Examination of the two linear changes of coordinates shows that the transformed f may have a sum norm (i.e. sum of the absolute values of its integer coefficients) of dn^n.

where d was the sum norm of the original polynomial (consider, for example, the use of a Horner type evaluation to actually do the changes of coordinates). We will assume that $L(dn^n) = L(d) + nL(n)$ is $O(nL(d))$, and so assume from now on that the length of the sum norm of our input polynomial is $O(nL(d))$. The cost of computing $pp(g)$ is less than the cost of the discriminant computation that we will count in Step (2).

10.2 TOPTYP Step (2)

In Step (1) of $AffinePlaneDecomp$, since $[0,1,0]$ is not on C, g has constant leading coefficient, hence $PROJ(\{g\}) = Discriminant(g)$. Discriminant computation is $Resultant(g, g')$, i.e. resultant of two bivariate polynomials of degree n or less and sum norm $n \cdot dn^n$. Note that the length of this new sum norm is still $O(nL(d))$. Computing the resultant of two bivariate polynomials of degree s and sum norm u takes time $O(s^5 L(u)^2)$ (Loos, 1982, p. 134), which gives us time $O(n^7 L(d)^2)$ altogether, and produces a polynomial of degree n^2 or less. If e is the sum norm of the discriminant, then $L(e) = O(n \cdot nL(d))$. The factorization of a univariate integral polynomial of degree s and sum norm u takes $O(s^{12} + s^9 L(u)^3)$ (Kaltofen, 1982, p. 111). Hence the factorization of $D(x)$ into irreducibles takes $O(n^{24} + n^{18} L(e)^3) = O(n^{24} + n^{24} L(d)^3) = O(n^{24} L(d)^3)$. For simplicity let us assume that the discriminant has only one irreducible factor; if it has more than one, that will complicate the analysis but will reduce the computing time. Root isolation applied to a squarefree univariate integral polynomial of degree s and sum norm u is $O(s^6 L(u)^2)$ (Collins & Loos, 1982, p. 93), so for the discriminant this gives us $O(n^{16} L(d)^2)$. The remaining actions of Step (1) of $AffinePlaneDecomp$ are not significant.

In Step (2) of $AffinePlaneDecomp$, let m be the number of roots of the discriminant $D(x)$ just computed; we have $m \leq n^2$. The dominant cost of this step is the root isolations of the polynomials $g(a_{2j}, y)$, $1 \leq j \leq m$, since in these cases a_{2j} is an algebraic, rather than a rational, number. The only bound presently available to us for root isolation of algebraic polynomials is that given for the Collins-Loos algorithm (Collins & Loos, 1982) by Rump (1975). We get a better bound assuming the use of an alternative, somewhat roundabout, algebraic polynomial root isolation algorithm. The alternative algorithm is to compute the "normal polynomial" $V(y) = Resultant(g(x, y), D(x))$ (whose roots include those of $g(a_{2j}, y)$), isolate its roots, and use $gsfd(g(a_{2j}, y))$ to select the intervals that contain a root of $g(a_{2j}, y)$. To compute the normal polynomial, we compute a resultant of one polynomial of degree n^2 or less and sum norm e (given our assumption that $D(x)$ has only one factor) and one of degree n or less and sum norm d; so this is $O(n^{10} L(e)^2) = O(n^{14} L(d)^2)$. The resultant is a polynomial of degree n^3 or less with sum norm v with $L(v) = O(nL(e))$, so root isolation takes $O(n^{27} L(d)^2)$ (we assume that $V(y)$ is squarefree). Let α denote the algebraic number a_{2i}, and let $g_\alpha(y) = g(a_{2j}, y)$. We must compute $gsfd(g_\alpha) = g_\alpha / gcd(g_\alpha, g'_\alpha)$. Evaluation of g at α takes no time; we just interpret $g(x, y)$ as a polynomial in y over $Q(\alpha)$. Suppose that we compute $gcd(g_\alpha, g'_\alpha)$ by a natural polynomial polynomial remainder sequence (Loos, 1982). At each of the $O(n)$ steps we have to do a division with remainder of two polynomials in $Q(\alpha)[y]$, each of which has degree n or less, and each of which has sum norm of $O(nL(n^2 d))$ (Loos, 1982, p. 133). Thus, since the degree of the minimal polynomial of α is n^2 or less, the $O(n)$ arithmetic operations in $Q(\alpha)$ we do at each step, each have a cost of $O((n^2)^3 n^2 L(nd)^2)$, thus our total cost for this gcd calculation is $O(n^{10} L(nd)^2)$. This surely dominates the cost of $g_\alpha / gcd(g_\alpha, g'_\alpha)$, and so we will take it to be the cost of the entire $gsfd$ computation. We next must evaluate the signs of the $gsfd$ at the endpoints of at most n^3 isolating intervals for roots of the polynomial $V(y)$ we computed above. Let u/v be one of these endpoints. Given that the degree of V_j is $O(n^3)$, by Horner's rule, this costs

$$n^3 L(u)\{n^3 L(u) + n^3 L(v) + nL(nd)\},$$

(Collins & Loos, 1982, p. 84), where $L(u)$ and $L(v)$ are each $O(n^3 L(n^3 d))$ (Collins & Loos (1982, p. 84). Thus the total time for the evaluation is $O(n^9 L(n^3 d)^2)$. The resulting element of $Q(\alpha)$ is represented as a polynomial with rational number coefficients, each of which has length that is $O(n^3 L(n^3 d))$. By Rump (1976), the sign of an algebraic number whose minimal polynomial has degree s, and such that u is the largest coefficient occurring either in the minimal polynomial or in the representing polynomial for that algebraic number, can be found in time $O(s^5 L(u)^3)$. Hence since in our case the degree of the minimal polynomial is n^2 or less, and the length of u is $O(n^3 L(n^3 d))$, we get a time of $O(n^{19} L(n^3 d)^3)$. Since we have $O(n^3)$ of these sign determinations to do, this gives us a total time of $O(n^{22} L(n^3 d)^3)$.

We have to do two more similar $gsfd$ calculations in this step for g_x and g_y, and evaluate those at the endpoints of the isolating intervals we have found to contain roots of $gsfd(g(\alpha, y))$, but as the cost of these computations is dominated by the cost of what we've already done for g, we take the cost of the Step (2) of $AffinePlaneDecomp$ for this i to be $O(n^{22} L(n^3 d)^3)$. Since $m = O(n^2)$, the total cost for Step (2) of $AffinePlaneDecomp$ is $O(n^{24} L(n^3 d)^3)$.

Now let us go on to Step (3) of $AffinePlaneDecomp$, and consider the cost of a single call to SSADJ2. Step (1) of SSADJ2 calls for a root isolation that we already did in Step (2) of $AffinePlaneDecomp$; we may assume that it is not repeated. In Step (2), we start knowing that (b_1, b_2) is an isolating interval for α as a root of its minimal polynomial $M(x)$; we must shrink (bisect) (b_1, b_2) until no $g(x, s_j)$ has a root in $[b_1, b_2]$, and (b_1, b_2) must still contain α. We can think of this as having to isolate the roots of product polynomials $M(x)\, g(x, s_j)$, for successive j. The cost of these isolations depends on the minimum root separation of $M(x)\, g(x, s_j)$ versus the minimum root separation of $M(x)$. For simplicity we will assume that the coefficients of these two polynomials have the same maximum size. Then it follows from Collins & Loos (1982, p. 84), that since the degree of $M(x)$ is $O(n^2)$, and the degree of $g(x, s_j)$ is $O(n)$, we have to do at most n bisections for each j, and since there are $O(n)$ successive values of j, we obtain a cost of $O(n^2)$ so far for step (2) of SSADJ2. Clearly this is not a significant cost, even given the fact that our discussion ignored the cost of the (rational number) arithmetic for each bisection. In the remaining steps of SSADJ2, we see that we have $O(n)$ calls to a root isolation algorithm for an integral polynomial of degree n. Let us count $O(n^6 L(d)^2)$ for each such call; this gives us $O(n^7 L(d)^2)$ total for one call to SSADJ2. SSADJ2 is executed $m = O(n^2)$ times, so altogether we have time $O(n^9 L(d)^2)$ for step (3) of $AffinePlaneDecomp$.

10.3 TOPTYP Step (3)

The only significant cost in Step (3) of TOPTYP is the root isolation of $f(1, y, 0)$, which is $O(n^6 L(d)^2)$.

10.4 TOPTYP Step (4)

Negligible cost.

10.5 TOPTYP Step (5)

We consider the computing time of the steps described in Section 7. First, note that the decomposition of the projective plane we construct has $O(n^3)$ cells, and $O(n^4)$ adjacencies. Thus the connectivity graph for our decomposition has $O(n^3)$ vertices and $O(n^4)$ edges. To construct the components of C involves constructing the connected components of connectivity graph, hence, if we use depth first search, then the time is $O(n^3 + n^4) = O(n^4)$ (Aho *et. al.*, 1974). As we have mentioned (cf. Abstract and Section 2), C has $O(n^2)$ components. For each component, we must determine if it is an oval (which we do with a connected components computation in a subgraph of the connectivity graph), and if so, do two Euler characteristic computations. Thus for each component of C, we have a cost of $O(n^4)$ for the connected components computation, and a cost of $O(n^3)$ for the Euler characteristic computations, hence a cost of $O(n^6)$ for all components of C. The cost of the steps we have described dominates the cost of determining the partial ordering of ovals. Hence the total time for this step is $O(n^6)$.

10.6 Summary

We see that altogether, the maximum computing time of TOPTYP is $O(n^{27}L(d)^3)$.

11 Acknowledgements

We are indebted to G. Brumfiel for the observation that Euler characteristic suffices to distinguish the interior of an oval from its exterior (we had originally contemplated a homology calculation). R.H. Bing was kind enough to provide us with a detailed account of some fundamental facts of the topology of plane curves. The second author would like to acknowledge helpful and inspiring conversations on the subject of this paper with the following people: G. Collins, E. Fadell, T.-C. Kuo, E. Mansfield.

12 References

Aho, A. V., Hopcroft, J., Ullman, J. (1974). *The Design and Analysis of Computer Algorithms.* Reading, Massachusetts: Addison-Wesley.

Arnon, D. S., McCallum, S. (1983). *A Polynomial-Time Algorithm for the Topological Type of a Real Algebraic Curve.* Tech. Rept. #454, Comp. Sci. Dept., Purdue Univ.

Arnon, D. S., McCallum, S. (1984). A polynomial-time algorithm for the topological type of a real algebraic curve – extended abstract. *Rocky Mountain J. Math.*, **14**, 849-852.

Arnon, D. S., Collins, G. E., McCallum, S. (1984a). Cylindrical algebraic decomposition I: the basic algorithm, *SIAM J. Comp.* **13/4**, 865–877.

Arnon, D. S., Collins, G. E., McCallum, S. (1984b). Cylindrical algebraic decomposition II: an adjacency algorithm for the plane, *SIAM J. Comp.* **13/4**, 878–889.

Arnon, D. S. (1988). A cluster-based cylindrical algebraic decomposition algorithm. *J. Symb. Comp.* **5**, (this issue).

Ben-Or, M., Kozen, D., Reif, J. (1986). The complexity of elementary algebra and geometry. *J. Comp. Sys. Sci.*, **32**, 251-264.

Collins, G. E. (1975). Quantifier elimination for real closed fields by cylindrical algebraic decomposition. Proceedings of the Second GI Conference on Automata Theory and Formal Languages. *Springer Lec. Notes Comp. Sci.* **33**, 515–532.

Collins, G. E. (1980). SAC-2 and ALDES now available. *ACM SIGSAM Bull.* **14**, 19.

Collins, G. E., Loos, R. G. K. (1982). Real zeros of polynomials. In (Buchberger, B., Loos, R., Collins, G. E., eds.) *Computer Algebra - Symbolic and Algebraic Computation* (Computing Supplementum 4), pp. 83-94. Vienna and New York: Springer-Verlag.

Fuks, D. (1974). Review of Polotovskii, G.M., An algorithm for determining the topological type of a structurally stable plane curve of even degree. *Math. Reviews 58*, No. 28000.

Gianni, P., Traverso, C. (1983). *Shape Determination for Real Curves and Surfaces.* Tech. Rept. #23, Math. Dept., Univ. Pisa.

Gudkov, D. A. (1974). The topology of real projective algebraic varieties. *Russian Math. Surveys* **29**, 1-79.

Kaltofen, E. (1982). Polynomial factorization. In (Buchberger, B., Loos, R., Collins, G. E., eds.) *Computer Algebra - Symbolic and Algebraic Computation* (Computing Supplementum 4), pp. 95-113. Vienna and New York: Springer-Verlag.

Kendig, K. (1978) *Elementary Algebraic Geometry.* New York, NY: Springer-Verlag.

Loos, R. (1982). Generalized polynomial remainder sequences. In (Buchberger, B., Loos, R., Collins, G. E., eds.) *Computer Algebra - Symbolic and Algebraic Computation* (Computing Supplementum 4), pp. 115-137. Vienna and New York: Springer-Verlag.

Massey W. S. (1978). *Homology and Cohomology Theory.* New York, NY: Marcel Dekker.

Milnor, J. (1965). *Topology from the Differentiable Point of View.* Charlottesville, VA: University Press of Virginia.

Moise, E. E. (1977). *Geometric Topology in Dimensions Two and Three.* New York, NY: Springer-Verlag.

Polotovskii, G. M. (1973). Algorithm for determining the topological type of a rough plane algebraic curve of even degree. In: Qualititative methods of the theory of differential equations and their applications, *Gor'kov Gos. Univ. Ucen. Zap. Vyp.* **187**, pp. 143-187 (Russian).

Ragsdale, V. (1906). On the arrangement of the real branches of plane algebraic curves. *Amer. J. Math.* **28**, 377-404.

Rump, S. (1975). *Isolierung der reellen Nullstellen algebraischer Polynome.* Bericht der Arbeitsgruppe Computer Algebra No. 10, FB Informatik, University of Kaiserslautern (German).

Rump, S. (1976). On the sign of a real algebraic number. *Proc. ACM SYMSAC '76.* Yorktown Heights, NY, (R.D. Jenks, ed.), 238-241.

Tarski, A. (1951). *A Decision Method for Elementary Algebra and Geometry.* Report R-109, second revised edn.. Santa Monica, CA: The Rand Corporation.

Walker, R. J. (1951). *Algebraic Curves.* Princeton, NJ: Princeton University Press.

Wilson, G. (1978). Hilbert's sixteenth problem. *Topology*, **17**, 53-73.