Daniel Plaumann
Universität Konstanz
Summer 2015

8. CURVES

8.1. Degree and genus

Let $X \subset \mathbb{P}^n$ be an irreducible curve of degree d. We know that the Hilbert polynomial p_X is a polynomial of degree 1 with rational coefficients. As such, it provides us with two numbers: We know that the leading coefficient of p_X is the degree of X. The meaning of the constant term $p_X(0)$ is less clear. Note first that this is always an integer (see Exercise 10.1).

Definition. The integer $g_a = 1 - p_X(0)$ is called the **arithmetic genus** of X. If X is smooth, then g_a is called the **geometric genus**, or simply the genus of X.

Thus if X is a curve of degree d and arithmetic genus g_a, then

$$p_X(m) = dm + 1 - g_a.$$

In general, the geometric genus of a curve is defined to be the arithmetic genus of its desingularization.

Examples 8.1. (1) Let X be a curve in $\mathbb{P}^2$ defined by an irreducible polynomial $F \in K[X, Y, Z]$ of degree d. We know from Example 7.4(2) that the Hilbert polynomial of X is $p_X(m) = d \cdot m - d(d-3)/2$, hence the arithmetic genus of X is

$$g_a = \frac{(d-1)(d-2)}{2} = \binom{d-1}{2}.$$

For the first few degrees, we find

d	1	2	3	4	5	6	7
g_a	0	0	1	3	6	10	15

(2) From the computation of the Hilbert function in Example 7.4(3), we know that the Hilbert polynomial of the twisted cubic C in $\mathbb{P}^3$ is $p_C(m) = 3m + 1$. Hence C is a curve of degree 3 and arithmetic genus 0. Since C is smooth, its arithmetic and geometric genus agree.

We have also seen in Exercise 2.4 that the projection $C' \subset \mathbb{P}^2$ of C from a point not on C is either a nodal or a cuspidal cubic. It follows that C' is a curve of arithmetic genus 1, but geometric genus 0, since C is its desingularization.

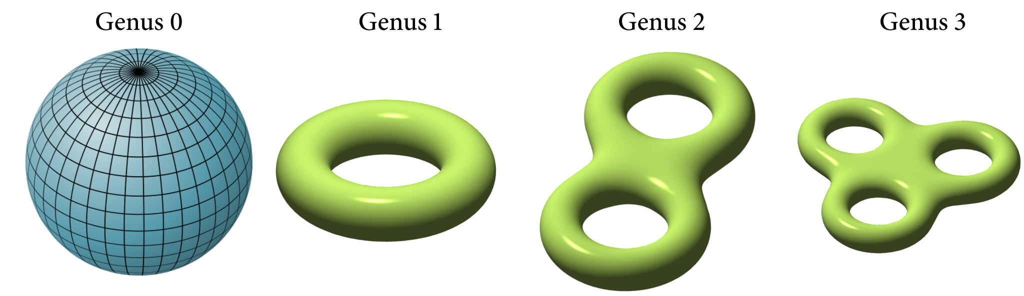
We see from (1) that not all positive integers occur as the arithmetic genus of a plane curve. However, the arithmetic genus of a curve is always greater or equal to 0 and for any $g \geqslant 0$ there exists a smooth curve of genus g (and therefore also a plane curve of geometric genus g, obtained by projecting a curve of genus g into the plane).

For the theory of algebraic curves, the genus is the most important invariant. This is so for a variety of different reasons, two of which we will briefly explain.

(1) **The genus of a smooth curve over $\mathbb{C}$ determines its topology.** A smooth curve over $\mathbb{C}$ leads a double life, first as an algebraic variety of dimension 1, second as a complex manifold of dimension 2. The complex manifold corresponding to the affine line $\mathbb{A}^1$ over $\mathbb{C}$ is the complex plane $\mathbb{C}$, topologically equal to $\mathbb{R}^2$. The projective line $\mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$ is the one-point compactification $\mathbb{C} \cup \{\infty\}$, also known as the Riemann sphere, topologically a two-dimensional sphere S^2. In general, we have the following amazing fact, the proof of which is outside our scope.

Theorem. *A smooth projective curve of genus g over $\mathbb{C}$ is an orientable surface of genus g, which means it is homeomorphic to the g-fold connected sum of the 2-torus.* ■

The genus is the number of 'holes' in the surface: A surface of genus 0 is sphere, of genus 1 a **torus**, genus 2 a **double torus**, and so on. For example, the complex points of a smooth plane curve of degree 3 form a torus, a fact which forms the basis for the analytic theory of elliptic curves.



(Source: Wikimedia Commons)

(2) **The genus of a smooth curve determines the behaviour of rational (or meromorphic) functions on the curve.** We know that a projective curve X does not admit any non-constant regular functions, in other words any morphism $X \to \mathbb{A}^1$ is necessarily constant. It follows that any non-constant rational function $X \dashrightarrow \mathbb{A}^1$ must have a **pole** somewhere, a point at which it is undefined. For example, a polynomial $f \in K[t]$ defines a morphism $\mathbb{A}^1 \to \mathbb{A}^1$, which we may interpret as a rational function $f\colon \mathbb{A}^1 \cup \{\infty\} = \mathbb{P}^1 \dashrightarrow \mathbb{A}^1$. This function has a pole of order $\deg(f)$ at the point ∞.

The general problem of locating the poles or realizing rational functions with prescribed poles is central to the theory of algebraic curves resp. of Riemann surfaces. Let $X \subset \mathbb{P}^n$ be an irreducible curve of degree d and assume that the hyperplane $\mathcal{V}(Z_0)$ intersects X transversely in the points $p_1, \dots, p_d \in X$. Given $G \in S(X)_m$, the fraction G/Z_0^m is a rational function on X, an element of the function field $K(X)$. As a function $X \dashrightarrow \mathbb{A}^1$, it is defined on $X \setminus \{p_1, \dots, p_d\}$, while in the points $p_1, \dots, p_d$ it has a pole. In fact, since Z_0 intersects X transversely, that pole is of order at most m; the pole order is equal to m if G does not vanish and p_i and lower than m if it does. In general, given points $p_1, \dots, p_k \in X$ and positive integers $m_1, \dots, m_k$, let

$$L(X, m_1p_1 + \cdots + m_kp_k) = \left\{ f \in K(X) : \begin{array}{l} f \text{ has a pole of order at most } m_i \text{ at } p_i \\ \text{for } i = 1, \dots, k \text{ and no other poles on } X. \end{array} \right\},$$

a linear subspace of $K(X)$. What we have just discussed is that there is a map

$$\iota\colon \begin{cases} S(X)_m & \to & L(X, mp_1 + \cdots + mp_d) \\ G & \mapsto & G/Z_0^m \end{cases}$$

It is injective, since $G/Z_0^m = H/Z_0^m$ on X for $G, H \in K[Z_0, \dots, Z_n]_m$ if and only if $Z_0^m \cdot (G - H) \in I(X)$. Since X is irreducible, $I(X)$ is prime and by assumption $Z_0 \notin I(X)$, hence $G - H \in I(X)$, in other words $G = H$ in $S(X)_m$. It follows that

$$\dim L(X, mp_1 + \cdots + mp_d) \geqslant h_X(m) = dm + 1 - g_a.$$

for sufficiently large m. This holds much more generally.

Theorem 8.2 (Riemann's inequality). *Let X be a smooth projective curve of genus g. Let $p_1, \dots, p_k \in X$ and $m_1, \dots, m_k \in \mathbb{Z}$. Then*

$$\dim L(X, m_1p_1 + \cdots + m_kp_k) \geqslant \sum_{i=1}^{k} m_i + 1 - g.$$

The dimensions of the spaces $L(X, m_1p_1 + \cdots + m_kp_k)$ can be seen as much more refined versions of the Hilbert function. (Note that the multiplicities $m_1, \dots, m_k$ are also allowed to be negative. In this case, having 'a pole of order at most m_i at p_i' is understood to mean 'vanishing to order at least $-m_i$ at p_i'.) The **Riemann-Roch theorem** in its full strength is an equation rather than an equality, i.e. it also describes the difference between the left and right hand side of Riemann's inequality. It also follows from this description that equality holds whenever $\sum_{i=1}^{k} m_i > 2g - 2$.

Recall that an irreducible curve X is called **rational** if it is birational to $\mathbb{P}^1$. This means that the function field $K(X)$ is isomorphic to the function field $K(\mathbb{P}^1) = K(t)$, the rational function field in one variable. Explicitly, for $X \subset \mathbb{P}^n$, this means that there is some dense open subset $U \subset X$ that can be **parametrized** through rational functions, i.e. there is a rational map

$$\varphi\colon \begin{cases} \mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\} & \dashrightarrow & X \\ t & \mapsto & [g_0(t)/h_0(t), \dots, g_n(t)/h_n(t)] \end{cases}$$

given by rational functions $g_i/h_i \in K(t)$, which is an isomorphism between non-empty open subsets of $\mathbb{P}^1$ and X, respectively.

It turns out that the geomeric genus exactly determines whether a curve is rational.

Theorem 8.3. *A smooth projective curve is rational if and only if its genus is* 0.

Sketch of proof. We know that the genus of $\mathbb{P}^1$ is 0, since it has Hilbert polynomial $p_{\mathbb{P}^1}(m) = m + 1$. That every other smooth rational curve has genus 0 is a consequence of the non-trivial fact that the geometric genus is preserved under birational maps. (This follows for instance from the Riemann-Roch theorem.) Conversely, suppose that X is a smooth curve of genus 0. Then for any point $p_0 \in X$, Riemann's inequality will give $\dim L(X; p_0) \geqslant 2$. Thus $L(X; p_0)$ contains a non-constant rational function f from $K(X)$. Consider the rational map $f\colon X \dashrightarrow \mathbb{P}^1$, $p \mapsto f(p)$. Since $f \in L(X; p_0)$, we know that $f^{-1}(\infty) = \{p_0\}$. Since p_0 is a simple pole of f, we can conclude from this that indeed the general fibre of f contains only a single point. By Thm. 5.3, this implies that f is a birational map. ∎

Example 8.4. We showed in Exercise 5.3 that the nodal cubic in $\mathbb{P}^2$ defined by $Y^2Z = X^3 - X^2Z$ has a rational parametrization given as the inverse of the projection from the point $[0,0,1]$. This curve has arithmetic genus 1 but geometric genus 0, since its desingularization is the twisted cubic in $\mathbb{P}^3$, as discussed above. If instead we take a smooth cubic in $\mathbb{P}^2$, like $X^3 + Y^3 = Z^3$ (if $\operatorname{char} K \neq 3$), it will have (geometric) genus 1 and therefore will not permit a rational parametrization, by the above theorem. This, however, is not so easy to prove directly. (Try for the example just given.)

We will neither prove nor use Riemann's inequality (moderately easy) or the Riemann-Roch theorem (more difficult). Instead, we will consider in the next section a somewhat related but more elementary geometric problem about plane curves.

8.2. The Cayley-Bacharach theorem

In this section, we discuss the Cayley-Bacharach theorem and its predecessors, a classical part of the theory of plane curves developed in the 19th century. Our exposition is based on the paper *Cayley-Bacharach Theorems and Conjectures* by Eisenbud, Green and Harris (*Bulletin of the American Mathematical Society*, **33**(5), 1996).

The story begins already in antiquity.

Theorem 8.5 (Pappus of Alexandria ~340 AD). *Let p, q, r and p', q', r' be two triples of collinear points in $\mathbb{P}^2$, all distinct and no four collinear. Then the three intersection points $\overline{pq'} \cap \overline{qp'}$, $\overline{pr'} \cap \overline{rp'}$ and $\overline{qr'} \cap \overline{rq'}$ are again collinear (see Figure 1).*

Of course, Pappus did not state his theorem in the projective plane. To obtain a true statement in the affine plane, one has to allow for a number of exceptional cases in which lines become parallel. Several different proofs are known. We will obtain it as a corollary from the following more general result proved by Michel Chasles at some point around 1840 (published 1865).

Theorem 8.6 (Chasles). *Let X_1, X_2 be two cubic curves in $\mathbb{P}^2$ meeting in exactly nine points $p_1, \ldots, p_9$. Then any cubic passing through $p_1, \ldots, p_8$ also passes through p_9.*

Chasles's theorem is often called the Cayley-Bacharach theorem, although that name should be reserved for a generalisation to curves of higher degree.

The proof will require some preparation. Note first that Pappus's theorem follows from that of Chasles: Let L and L' be the two lines containing p, q, r and p', q', r', respectively. Let X_1 be the cubic obtained as the union of $\overline{pq'}$, $\overline{qr'}$ and $\overline{rp'}$ and X_2 the union of $\overline{pr'}$, $\overline{qp'}$, $\overline{rq'}$. The nine intersection points of X_1 and X_2 are

$$\begin{aligned} p &= \overline{pq'} \cap \overline{pr'}, & q &= \overline{qr'} \cap \overline{qp'}, & r &= \overline{rp'} \cap \overline{rq'} \\ p' &= \overline{rp'} \cap \overline{qp'}, & q' &= \overline{pq'} \cap \overline{rq'}, & r' &= \overline{qr'} \cap \overline{pr'} \\ a &= \overline{qr'} \cap \overline{rq'}, & b &= \overline{rp'} \cap \overline{pr'}, & c &= \overline{pq'} \cap \overline{qp'}. \end{aligned}$$

Now apply Chasles's theorem to the cubic X obtained as the union of L, L' and $\overline{ab}$. Then X passes through $p, p', q, q', r, r', a, b$, hence it also passes through c. The hypothesis implies that c cannot lie on L or L', so it must lie on $\overline{ab}$, as claimed.

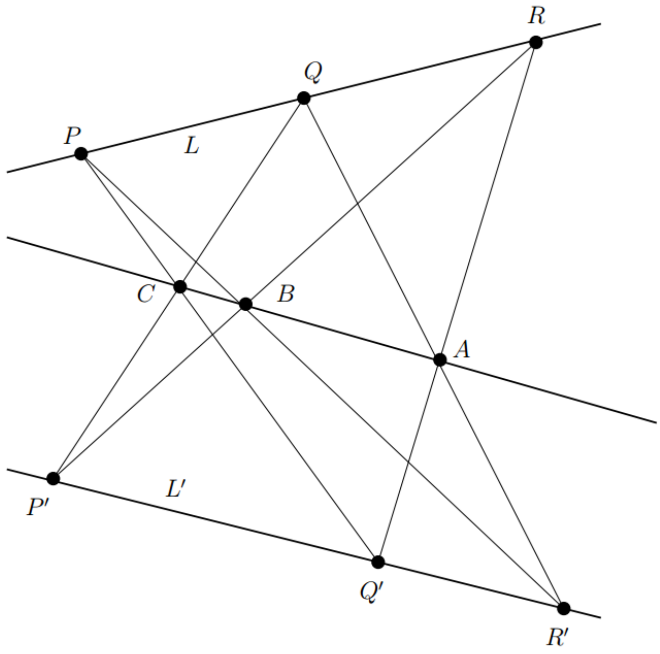


FIGURE 1. Pappus's theorem
(Source: math.stackexchange.com - original source unknown)

Before we prove Chasles's theorem, let us try to better understand its meaning. Let F_1 and F_2 be two non-zero homogeneous polynomials of degree 3, $X_1 = \mathcal{V}(F_1)$, $X_2 = \mathcal{V}(F_2)$. Put $V = K[X,Y,Z]_3$ and let

$$\mu_i : V \to K,\ F \mapsto F(p_i)$$
$$H_i = \{F \in V : F(p_i) = 0\} = \ker(\mu_i)$$

for $i \in \{1,\dots,9\}$. The function μ_i is a linear functional, called the **point evaluation** at p_i. Its kernel H_i is a hyperplane in V. The intersection of nine general hyperplanes in V is 1-dimensional. However, $H_1 \cap \dots \cap H_9$ contains the two cubics F_1 and F_2 and has therefore dimension at least 2. It follows that $\mu_1,\dots,\mu_9 \in V^*$ span a subspace of dimension at most 8 in V^* and are therefore linearly dependent. Let

$$\alpha_1\mu_1 + \dots + \alpha_9\mu_9 = 0$$

be a non-trivial linear relation with $a_j \neq 0$, then we can write $\mu_j = (-1/\alpha_j)\sum_{i\neq j}\alpha_i\mu_i$. It follows that any polynomial in V vanishing at the eight points $\{p_i : i \neq j\}$ also vanishes at p_j.

Does this bit of linear algebra prove the theorem? Not quite, since it gives us no control over the question which of the coefficients $\alpha_1,\dots,\alpha_9$ is non-zero. It only says that *one* of them must be non-zero. The statement of Chasles's theorem is equivalent to the statement that *any* one of $\mu_1,\dots,\mu_9$ is a linear combination of the other eight, in other words that any eight span the same subspace of V^*. Since the two cubics X_1 and X_2 may be very degenerate (as for instance in the proof of Pappus's theorem), it is by no means clear that any one of the nine points behaves like any other.

We will prove Chasles's theorem by analysing more precisely the Hilbert function of finitely many points. Remember first our solution of Exercise 8.1 (Prop. 7.2).

Lemma 8.7. *Let Γ be a set of k points in $\mathbb{P}^n$ and let $d \geqslant 0$. The following are equivalent:*

(1) The Hilbert function h_Γ satisfies $h_\Gamma(d) = k$.

(2) The point evaluations $\{\mu_p : p \in \Gamma\} \subset K[Z_0, \dots, Z_n]_d^*$ are linearly independent, where $\mu_p : K[Z_0, \dots, Z_n]_d \to K$, $F \mapsto F(p)$.

(3) For each $p \in \Gamma$ there exists $F \in K[Z_0, \dots, Z_n]_d$ vanishing at all the points in $\Gamma \setminus \{p\}$ but not at p.

Proof. Write $V = K[Z_0, \dots, Z_n]_d$, $\Gamma = \{p_1, \dots, p_k\}$, $\mu_i = \mu_{p_i}$. Consider the map

$$\varphi : \begin{cases} V & \to & K^k \\ F & \mapsto & (\mu_1(F), \dots, \mu_k(F)) \end{cases}.$$

Then $I(\Gamma)_d$ is the kernel of φ, so that $h_\Gamma(d)$ is the dimension of the image. Hence $h_\Gamma(d) = k$ implies that φ is surjective. This means that for each $i \in \{1, \dots, k\}$, there is $F_i \in V$ with $\varphi(F_i) = e_i$, which shows (1)⇒(3). Also, if $\sum_{i=1}^k a_i \mu_i$ is a linear relation, evaluating at F_i shows $a_i = 0$, showing (1)⇒(2). Conversely, (3) implies that all unit vectors $e_1, \dots, e_k$ are in the image of φ, so (3)⇒(1). Finally, if φ is not surjective, its image is contained in a hyperplane in K^k, which shows that $\mu_1, \dots, \mu_k$ are linearly dependent, hence (2)⇒(1). ■

The following proposition will be the key to the proof of Chasles's theorem. While we only need a special case ($d = 3$ and $k = 8$), the statement and its rather intricate proof actually become clearer when made for any number of points.

Proposition 8.8. *Fix $d \geqslant 0$ and let $\Gamma \subset \mathbb{P}^2$ be a set of k points where $k \leqslant 2d + 2$. Then the Hilbert function h_Γ satisfies $h_\Gamma(d) < k$ if and only if one of the following occurs:*

(i) Γ contains $d + 2$ collinear points.

(ii) $k = 2d + 2$ and Γ is contained in a conic.

Proof. Again write $V_d = K[X, Y, Z]_d$, $\Gamma = \{p_1, \dots, p_k\}$ and let $\mu_i : V_d \to K$ be the point evaluation at p_i.

First suppose that (i) holds and assume after relabelling that $p_1, \dots, p_{d+2}$ are contained in a line $\mathcal{V}(L)$. Then any form in V_d that vanishes on Γ must vanish on $\mathcal{V}(L)$ and is therefore divisble by L. We conclude

$$\{F \in V_d : F(p_1) = \cdots = F(d_{d+2}) = 0\} \subset L \cdot V_{d-1}.$$

Since the co-dimension of V_{d-1} in V_d is $\binom{d+2}{2} - \binom{d+1}{2} = d + 1$, the point evaluations $\mu_1, \dots, \mu_{d+2}$ span a subspace of dimension at most $d + 1$ in V_d^*. Thus $\mu_1, \dots, \mu_k$ span a subspace of dimension at most $d + 1 + (k - d - 2) = k - 1$ and are therefore linearly dependent. Similarly, if (ii) holds, the existence of a conic $\mathcal{V}(Q)$ containing Γ implies that $\mu_1, \dots, \mu_k$ span a subspace of dimension at most $\binom{d+2}{2} - \binom{d}{2} = 2d + 1$, so that $h_\Gamma(d) \leqslant 2d + 1 < 2d + 2 = k$.

The converse direction is harder. We use nested induction, first on d and then on k. For $d = 1$, the statement is clear: If $k \leqslant 4$ and $h_\Gamma(1) < k$, then we must have $k = 4$ (hence (ii)) or $k = 3$ and p_1, p_2, p_3 are collinear (c.f. Examples 7.1).

Now let $d \geqslant 2$. Suppose, for the sake of contradiction, that $h_\Gamma(d) < k$, but neither (i) nor (ii) hold. From Prop. 7.2 (a.k.a. Exercise 8.1) we know that $h_\Gamma(d) = k$ as long as $k \leqslant d + 1$. So we must have $k > d + 1$. Keeping d fixed and applying the induction hypothesis for k, we may conclude that $h_{\Gamma'}(d) = k - 1$ holds for any subset $\Gamma' \subset \Gamma$ of

$k-1$ points. By Lemma 8.7, the point evaluations $\mu_1, \dots, \mu_k$ are linearly dependent, but any $k-1$ of them are linearly independent. Therefore, any $G \in V_d$ vanishing at all but one point of Γ vanishes on all of Γ. We now distinguish two cases.

(a) Suppose Γ contains three collinear points. Let $\mathcal{V}(L)$ be a line containing the maximal number $m \geqslant 3$ of collinear points in Γ. Since (i) fails, we must have $m \leqslant d+1$. Let $\Gamma' = \Gamma \setminus \mathcal{V}(L)$ be the $k-m$ remaining points. We apply the induction hypothesis for $(k-m, d-1)$ to Γ'. We cannot have $k-m = 2(d-1)+2$, since $m \geqslant 3$, so (ii) cannot hold for Γ' in degree $d-1$. If Γ' contains $d+1$ points on a line, then we must have $m \geqslant d+1$ by the maximality of $\mathcal{V}(L)$, hence $k = 2d+2$ and $m = d+1$. But then all points of Γ' lie on a line $\mathcal{V}(L')$, so that Γ is contained in the conic $\mathcal{V}(LL')$ and (ii) holds for Γ, a contradiction. The induction hypothesis therefore yields

$$h_{\Gamma'}(d-1) = k-m.$$

Fix any $p \in \Gamma'$. Applying Lemma 8.7, we find $G \in V_{d-1}$ such that $G(p) \neq 0$ and $G(q) = 0$ for all $q \neq p$ in Γ'. Then $L \cdot G$ vanishes on all of Γ except at p, a contradiction.

(b) Suppose that no three points in Γ are collinear. Let p_1, p_2, p_3 be any three points in Γ and let $\Gamma'' = \Gamma \setminus \{p_1, p_2, p_3\}$. If $h_{\Gamma'' \cup \{p_i\}}(d-1) = k-2$ for some $i \in \{1,2,3\}$, we are done by the same argument as in (a). Otherwise, since $\Gamma'' \cup \{p_i\}$ cannot contain $d+1$ points on a line, the induction hypothesis implies $k = 2d+2$ and each of the sets $\Gamma'' \cup \{p_i\}$ is contained in a conic C_i. If $d = 2$, then Γ consists of six points and $h_\Gamma(2) < 6$ implies that Γ is contained in a conic, so we are done. If $d \geqslant 3$, then Γ'' contains at least 5 points, no three of which are collinear. So there can be only one conic containing Γ'', hence $C_1 = C_2 = C_3$. But then this conic contains Γ, hence (ii) holds. This completes the proof. ■

Proof of Chasles's theorem (8.6). Let X_1, X_2 be two cubics intersecting in exactly nine points $p_1, \dots, p_9$. We have seen that the point evaluations $\mu_1, \dots, \mu_9$ span a space of dimension at most 8. We must show that any eight, say $\mu_1, \dots, \mu_8$, span the same space. In fact, we prove the stronger statement that $\mu_1, \dots, \mu_8$ are linearly independent. Applying Prop. 8.8 with $\Gamma = \{p_1, \dots, p_8\}$, $k = 8$, $d = 3$, we see that $h_\Gamma(3) < 8$ would imply that (i) four points in Γ lie on a line ℓ or that (ii) $\{p_1, \dots, p_8\}$ lie on a conic C. If (i), both X_1 and X_2 would contain ℓ, contradicting $|X_1 \cap X_2| < \infty$. If (ii) and X_1 contains no component of C, then $|C \cap X_1| \leqslant 6$ by Bézout's theorem (7.13), and the same for X_2. Hence C has a common component with X_1 and with X_2. Since no component of C can be contained in $X_1 \cap X_2$, the only possibilty left to consider is that C is the union of two distinct lines ℓ_1 and ℓ_2 with $\ell_1 \subset X_1$ and $\ell_2 \subset X_2$. But since $\ell_1 \cup \ell_2$ contains Γ, this is easily seen to be impossible. ■

Chasles's theorem admits the following generalisation to curves of higher degree.

Theorem 8.9 (Cayley-Bacharach theorem). *Let X_1, X_2 be two curves in $\mathbb{P}^2$ of degree d and e, respectively, such that $\Gamma = X_1 \cap X_2$ consists of exactly $d \cdot e$ points. If X is any curve of degree $d+e-3$ containing all but one point of Γ, then X contains Γ.*

More generally, suppose that Γ is the disjoint union of two subsets Γ' and Γ''. Then

$$h_\Gamma(m) = h_{\Gamma'}(m) - h_{\Gamma''}(d+e-3-m) + |\Gamma''|.$$

holds for any $m \leqslant d+e-3$. ■

To see that the more general statement about the Hilbert function implies the first, let Γ'' be a single point and put $m = d + e - 3$. Since $h_{\Gamma''}(0) = 1$, the equation becomes $h_\Gamma(d + e - 3) = h_{\Gamma'}(d + e - 3)$, which says exactly that any form of degree $d + e - 3$ vanishing on Γ' vanishes on all of Γ.

One might think that since we proved Prop. 8.8 for any number of points, at least the first form of this result might follow just as easily as Chasles's theorem. However, upon inspection, we see that we would need $de - 1 \leqslant 2d + 2e - 4$, which happens only if $d = e = 3$ (or if one of d, e is at most 2, but then case (ii) in Prop. 8.8 will not lead to a contradiction).

A modern proof of Thm. 8.9 uses the Riemann-Roch theorem and another fundamental result about plane curves known as the Brill-Noether residue theorem. (It is worth pointing out that, with all this machinery, the proof at last becomes fairly easy, certainly much shorter than the proof of Chasles's theorem given above.)

Remark. The article of Eisenbud, Green and Harris, on which our exposition is based, presents several further generalisations of the Cayley-Bacharach theorem, also in higher dimensions, and some open problems. It also contains many interesting historic remarks. One concerns the role of the great Arthur Cayley. His contribution to the 'Cayley-Bacharach theorem' consisted in an even stronger version that he published early in his career, which however turned out to be completely wrong (with hindsight, in a fairly obvious way). This shows how even great mathematicians like Cayley, rightly famous for his numerous important contributions to algebra and geometry, can make basic mistakes in print.

Less known is Isaak Bacharach, who first proved the Cayley-Bacharach theorem. A German mathematician at the University of Erlangen, he was deported for being Jewish by the Nazis and died at the Theresienstadt concentration camp in 1942.

8.3. Real curves and Harnack's theorem

In this section, we look at another very classical part of the theory of plane curves: the topology of curves in the real projective plane.

First, we consider the situation over the complex numbers. Affine space $\mathbb{A}^n = \mathbb{C}^n$ carries not only the Zariski topology but also the usual topology of $\mathbb{R}^{2n}$, which we refer to as the strong or **Euclidean topology** on $\mathbb{A}^n$. It is finer than the Zariski topology, i.e. it has more open and closed sets. Complex projective space $\mathbb{P}^n$ also has a Euclidean topology which can be defined in two ways: We can define it locally and say that a set $U \subset \mathbb{P}^n$ is open if and only if $U \cap U_i$ is open in the affine space $U_i = \{[Z] \in \mathbb{P}^n : Z_i \neq 0\}$ for all $i = 0, \dots, n$. Or we can define it on $\mathbb{P}^n$ as the quotient topology obtained from the Euclidean topology on $\mathbb{A}^{n+1}$ via the quotient map $\mathbb{A}^{n+1} \setminus \{0\} \to \mathbb{P}^n$. Finally, any subvariety V of $\mathbb{A}^n$ or $\mathbb{P}^n$ also carries the Euclidean topology as the subspace topology induced on V, in which a subset $U \subset V$ is open if and only if there exists an open subset U' in the ambient space such that $U = U' \cap V$.

The Zariski topology is a very coarse topology with many irreducible subsets. By contrast, the only irreducible sets in the Euclidean topology are single points. In particular, an irreducible variety of positive dimension is not irreducible in the Euclidean topology. However, it is still connected in the Euclidean topology. We will show this in the case of curves.

Theorem 8.10. *Any irreducible curve over $\mathbb{C}$ is connected in the Euclidean topology.*

Sketch of proof. Since connectedness is not affected by adding or deleting finitely many points from X, we may assume that $X \subset \mathbb{P}^n$ is smooth and projective.

Suppose that X is not connected and let $X = M_1 \cup M_2$ be a decomposition into two disjoint closed subsets of X. At least one of M_1, M_2 must be infinite, say M_1. Also, since X is compact, it can have at most finitely many connected components, so we may assume that M_1 is connected. Let $p_0 \in M_2$ be any point. There exists a non-constant rational function $f \in \mathbb{C}(X)$ defined on $X \setminus \{p_0\}$. This directly follows from Riemann's inequality (Thm. 8.2) applied to $L(X, mp_0)$ for m sufficiently large. Hence $f\colon M_1 \to \mathbb{C}$ is defined everywhere and continuous. Let $q \in M_1$ be a point where $|f|$ attains its maximum on M_1. Since f is analytic in some neighbourhood U of q in $\mathbb{P}^n$, it attains its maximum only on the boundary of U unless f is constant on U, by the maximum modulus principle. But q is an interior point of U, so f must be constant on U, say $f|_U = \alpha$. Then the function $f - \alpha \in \mathbb{C}(X)$ vanishes identically on U. This means that $f - \alpha$ has infinitely many zeros on X, but is non-constant, since it has a pole at p_0. This contradiction shows the claim. ∎

The same is true in complete generality.

Theorem 8.11. *Any irreducible quasi-projective variety over $\mathbb{C}$ is connected in the Euclidean topology.* ∎

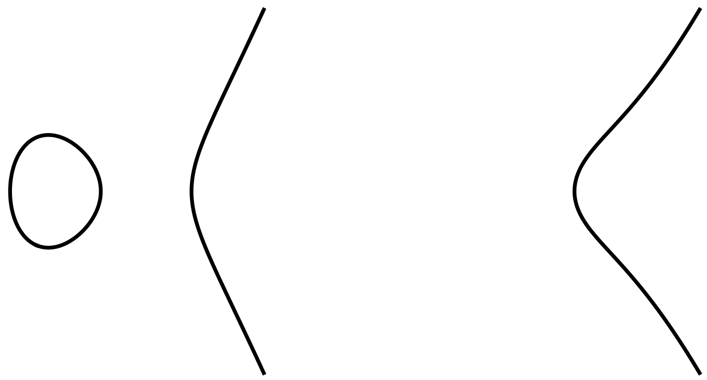
We do not give a proof here, but we mention that this can be deduced from Thm. 8.10 by an induction argument on the dimension of X (see Shafarevich, *Basic Algebraic Geometry II*, Ch. VII, §2.2).

Over the real numbers, the situation is completely different. Let $F \in \mathbb{R}[X, Y, Z]$ be an irreducible homogeneous polynomial of degree d and let $C = \mathcal{V}(F)$. We say that C is a **real curve**. It is usually not true that the set of real points

$$C(\mathbb{R}) = \mathcal{V}_{\mathbb{R}}(F) = \{p \in \mathbb{P}^2(\mathbb{R}) : F(p) = 0\}$$

is connected.

Example 8.12. Consider the cubic curve $C = \mathcal{V}((Y^2Z - (X-Z)(X-2Z)(X-3Z))$. Looking at the affine part where $Z = 1$, we see that $C(\mathbb{R})$ has two connected components. On the other hand, if $C = \mathcal{V}(Y^2Z - X(X^2+Z^2))$, then $C(\mathbb{R})$ is connected.



The question we want to answer in this section is: How many connected components can a real curve of degree d in the projective plane have?

Before we discuss this, we need to know a little about the individual components. First note that the real projective plane $\mathbb{P}^2(\mathbb{R})$ is a compact two-dimensional manifold. The manifold structure is given by the usual affine cover and compactness follows from the fact that $\mathbb{P}^2(\mathbb{R})$ is a quotient of the compact set S^2.

Now if $C \subset \mathbb{P}^2$ is a smooth curve, then $C(\mathbb{R})$ is a compact one-dimensional submanifold (by the implicit function theorem). It therefore has finitely many connected components and each component is a compact connected one-dimensional manifold. There is only one such manifold up to homeomorphism (or even diffeomorphism), namely S^1, the circle. (If you attended my topology class last semester, you have seen a proof of this fact.) Thus all connected components of $C(\mathbb{R})$ are homeomorphic to S^1. There is one important subtlety, however. If $M \subset \mathbb{R}^2$ is an embedded circle, then $\mathbb{R}^2 \setminus M$ has two connected components, one bounded, the other unbounded. This fact is known as the **Jordan curve theorem**.

Not so in the real projective plane: There are two topologically distinct embeddings of a circle into $\mathbb{P}^2(\mathbb{R})$. Consider the ellipse $C = \mathcal{V}(X^2+Y^2-Z^2)$ and a line $L = \mathcal{V}(X)$. Both are smooth and the real points are non-empty and connected, so $C(\mathbb{R})$ and $L(\mathbb{R})$ are both homeomorphic to S^1. But the difference is that $\mathbb{P}^2 \setminus L(\mathbb{R})$ is $\mathbb{A}^2(\mathbb{R}) = \mathbb{R}^2$, while $\mathbb{P}^2(\mathbb{R}) \setminus C(\mathbb{R})$ has two connected components, the 'outside' and the 'inside' of the ellipse. (To see this, note that the line $\mathcal{V}(Z)$ does not meet C in a real point. Thus $C(\mathbb{R})$ is contained in the affine plane $U(\mathbb{R}) = \mathbb{R}^2$, where $U = \mathbb{P}^2 \setminus \mathcal{V}(Z)$. Now $(C \cap U)(\mathbb{R})$ is the unit circle, defined by $x^2 + y^2 - 1 = 0$, hence its complement in $\mathbb{R}^2$ has two connected components. The bounded component, the interior of the unit disc, is then also a component of $C(\mathbb{R})$.) In other words, the topology of $C(\mathbb{R})$ and $L(\mathbb{R})$ is the same, but the embedding is different, because the topology of the complement $\mathbb{P}^2(\mathbb{R}) \setminus C(\mathbb{R})$ and $\mathbb{P}^2(\mathbb{R}) \setminus L(\mathbb{R})$ is different.

This leads to the following definition: If $M \subset \mathbb{P}^2(\mathbb{R})$ is homeomorphic to S^1, then M is called an **oval** if $\mathbb{P}^2(\mathbb{R})$ has two connected components. Otherwise, M is called a **pseudo-line**. If M is an oval, then $\mathbb{P}^2(\mathbb{R}) \setminus M$ is the union of the **interior** of M, which is homeomorphic to an open disc, and the **exterior** of M, which is homeomorphic to a Moebius strip. If L is a pseudo-line, then $\mathbb{P}^2(\mathbb{R}) \setminus U$ is homeomorphic to $\mathbb{R}^2$ (and thus also to an open disc). The most important difference for us is the following.

Lemma 8.13. *Let C_1 and C_2 be two real curves without common components and assume that C_1 and C_2 intersect transversely.*

1. *If $M \subset C_1(\mathbb{R})$ is an oval, then $M \cap C_2(\mathbb{R})$ consists of an even number of points.*
2. *If $L_1 \subset C_1(\mathbb{R})$ and $L_2 \subset C_2(\mathbb{R})$ are two pseudo-lines, then $L_1 \cap L_2$ consists of an odd number of points.*

If C_1 and C_2 do not intersect transversely, the statements remain true if the intersection points are counted with multiplicities.

Sketch of proof. We do not quite have the topological tools to give a full proof, but we sketch the proof of the first statement. (The second is somewhat more difficult.)

Let U, V be the components of $\mathbb{P}^2(\mathbb{R}) \setminus M$. Let N be a connected component of $C_2(\mathbb{R})$. Since N is homeomorphic to S^1, there is a parametrization $\varphi\colon [0,1] \to N$ with $\varphi(0) = \varphi(1)$ which is continuous and injective on $[0,1)$. If $M \cap N = \varnothing$, there is nothing to show. So suppose $M \cap N \neq \varnothing$ and let $t_1 < \cdots < t_n \in [0,1)$ be such that $\varphi(t_1), \ldots, \varphi(t_n) \in M \cap N$ are all the intersection points of M and N. We may assume that $\varphi(0) \notin M$, so that $t_1 > 0$, otherwise we reparametrize. Now since the intersection of M and N is transversal, N has to pass from the exterior into the interior or vice-versa in every intersection point. In other words, for each i there is $\varepsilon > 0$ such that $\varphi(t_i - \varepsilon, t_i) \subset U$ and $\varphi(t_i, t_i + \varepsilon) \in V$ or the other way around. It follows that if $\varphi(0) \in U$, then $\varphi(t_i, t_{i+1}) \in U$ if i is even and $\varphi(t_i, t_{i+1}) \in V$ if i is odd. Now $\varphi(0, t_1), \varphi(t_n, 1) \in U$, so n must be even, as claimed. (If M is contained in an affine plane $U_i(\mathbb{R}) \cong \mathbb{R}^2$, there is a simpler way of proving this; see Exercise 12.1.) ∎

Theorem 8.14. *Let C be a smooth real curve of degree d in $\mathbb{P}^2$. If d is even, then $C(\mathbb{R})$ is a finite union of ovals. If C is odd, then $C(\mathbb{R})$ is a finite union of ovals and exactly one pseudo-line.*

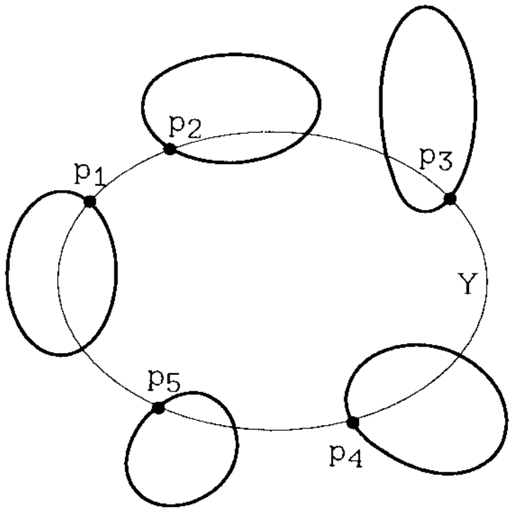
Proof. Since any two pseudo-lines intersect, there cannot be more than one among the connected components of $C(\mathbb{R})$. The degree coincides with the number of intersection points of C with a general line. If L is a real line, the number of non-real intersection points is even, since such points come in complex-conjugate pairs. On the other hand, the predecing lemma implies that the number of intersection points of $L(\mathbb{R})$ and $C(\mathbb{R})$ is even if and only if $C(\mathbb{R})$ contains no pseudo-line. This proves the claim. ∎

Theorem 8.15 (Harnack)**.** *Let C be a smooth real curve of degree d in $\mathbb{P}^2$. Then $C(\mathbb{R})$ has at most $g+1$ connected components, where $g = (d-1)(d-2)/2$ is the genus of C.*

Proof. We know that lines ($d = 1$) and conics ($d = 2$) are connected. So we assume $d \geqslant 3$ and suppose for the sake of contradiction that $C(\mathbb{R})$ has at least $g+2$ connected components. Then $C(\mathbb{R})$ contains at least $g+1$ ovals $M_1, \ldots, M_{g+1}$ and one more connected component N (which may be either an oval or a pseudo-line). Since the space of polynomials of degree $d-2$ in 3 variables has dimension $\binom{d}{2} = d(d-1)/2$, we can pick any $d(d-1)/2 - 1$ points in $\mathbb{P}^2$ and always find a curve of degree $d-2$ passing through these points. So pick $p_1, \ldots, p_{g+1}$ with $p_i \in M_i$ and the remaining $d(d-1)/2 - 1 - (g+1) = d-3$ points $q_1, \ldots, q_{d-3}$ on N. Now let C' be a real curve of degree $d-2$ passing through all these points, then $C'(\mathbb{R})$ intersects each oval M_i, $i = 1, \ldots, g+1$ at least twice, so that we find a total of

$$d(d-1)/2 - 1 + g + 1 = d(d-2) + 1$$

intersection points of C and C' (counted with multiplicity if the intersection is not transversal). This contradicts Bézout's theorem (Thm 7.13), which says there can be at most $d(d-2)$ intersection points.



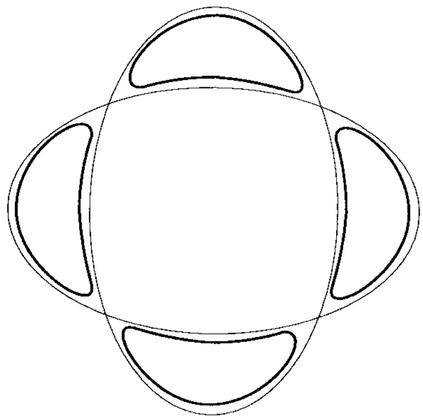
Source: [Ha], p. 248; Y is $\mathcal{V}(G)$ in the text.

■

Remark 8.16. If C is not smooth, the statement of Harnack's theorem remains true if we take $g = (d-1)(d-2)/2$ to be the arithmetic genus.

Harnack's theorem leads naturally to many further questions: First of all, is the bound in Harnack's theorem sharp, i.e. does there always exist a smooth curve of degree d with the maximal number $g+1$ of components? The answer is yes and a construction was already given by Harnack himself. The construction in general is quite involved and we do not discuss it.

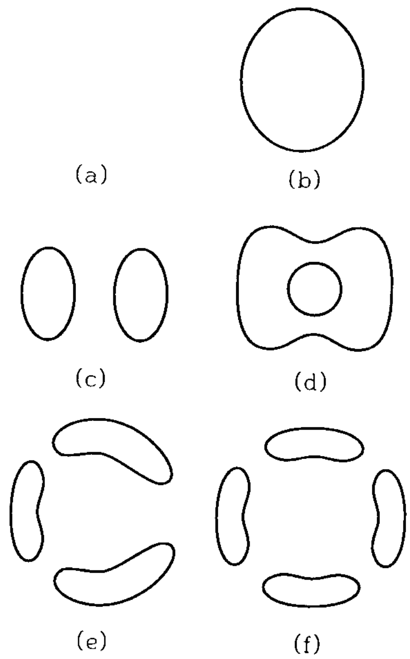
Example 8.17. The first case which is not completely trivial is $d = 4$: By Harnack's theorem, a smooth curve of degree 4 can have at most 4 ovals. To find a smooth quartic with 4 ovals, we can use the following trick: Start with two ellipses intersecting in four real points. For example, we may take $G = (X^2 + 4Y^2 - Z^2)(4X^2 + Y^2 - Z^2)$. Of course, $\mathcal{V}(G)$ is not smooth. But it can be smoothened out by a small **perturbation**: Let $H \in \mathbb{R}[X,Y,Z]_4$ be any other quartic and consider the curve $C = \mathcal{V}(G + \varepsilon H)$. For general H and ε, this curve will be smooth (since the set of singular quartics, the discriminant, is a proper subvariety of the space $\mathbb{R}[X,Y,Z]_4$). On the other hand, for ε small enough, the real picture $C(\mathbb{R})$ will be 'close' to that of the two ellipses. This will result in the desired curve with four ovals, as in the following picture:



Source: [Ha], p. 249

Another question concerns the relative position of the ovals. (The precise formulation of the question is to determine the homotopy type of the embedding $C(\mathbb{R}) \hookrightarrow \mathbb{P}^2(\mathbb{R})$.) This amounts to deciding what are the possible **nestings** of ovals.

Example 8.18. Let us try to understand this for quartics (c.f. Exercise 12.2.)
(a) Clearly, there are smooth quartics with no real points, for example we may take $\mathcal{V}(X^4 + Y^4 + Z^4)$.
(b) Finding a smooth quartic with just one oval is also easy, for example we may take $\mathcal{V}(X^4 + Y^4 - Z^4)$.
(c), (d) To find a quartic with two ovals, we can start with two disjoint ellipses and apply the perturbation trick above. Note that there are two possibilities: One ellipse can be *nested* inside the other, or not.
(e) A quartic with three ovals can also be found with the perturbation trick.
(f) A quartic with four ovals was constructed above.
Note that if a quartic has more than two ovals, there can be no nesting. For otherwise we could find six collinear points on those ovals, which is impossible for a curve of degree 4.



Source: [Ha], p. 249

In general, the question for the relative position of the ovals in any degree has turned out to be very difficult. It was posed by Hilbert as part of the sixteenth problem on his famous list of open mathematical problems presented at the 1900 International Congress of Mathematicians. For example, for $d = 6$ it has been shown that for a curve with the maximal number of ovals (which is $5 \cdot 4/2 + 1 = 11$), there are only three possible configurations. In general, the problem remains open.