ALGEBRAIC CURVES

B3b course 2009

Nigel Hitchin

hitchin@maths.ox.ac.uk

These notes trace a path through material which is covered in more detail in the book for the course which is:

F Kirwan, *Complex Algebraic Curves*, LMS Student Texts 23, Cambridge 1992, Chapters 2–6, £26.99

Contents

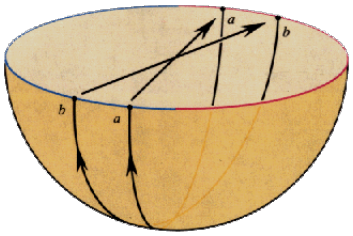
1 Projective spaces

1.1 Basic definitions

Definition 1 *Let V be a vector space. The projective space $P(V)$ of V is the set of 1-dimensional vector subspaces of V.*

Definition 2 *If the vector space V has dimension $n+1$, then $P(V)$ is a projective space of dimension n. A 1-dimensional projective space is called a projective line, and a 2-dimensional one a projective plane.*

For most of the course, the field F of scalars for our vector spaces will be the complex numbers $\mathbf{C}$. Our intuition is best served, however, by thinking of the real case. So the projective space of $\mathbf{R}^3$ is the set of lines through the origin. Each such line intersects the unit sphere $S^2 = \{x \in \mathbf{R}^3 : \sum_i x_i^2 = 1\}$ in two points $\pm u$, so from this point of view $P(\mathbf{R}^3)$ is S^2 with antipodal points identified. Since each line intersects the lower hemisphere, we could equally remove the upper hemisphere and then identify opposite points on the equatorial sphere.



In the **B3a** *Geometry of Surfaces* course this is the way we think of the projective plane, but it is less appropriate for an algebraic geometry course. Still, it does explain why we should think of $P(\mathbf{R}^{n+1})$ as n-dimensional. In what follows we shall write $P(F^{n+1})$ as $\mathbf{P}_n$ (usually for $F = \mathbf{C}$) to make this more plain.

A better approach for our purposes is the notion of a representative vector for a point of $P(V)$. Any 1-dimensional subspace of V is the set of multiples of a non-zero vector $v \in V$. We then say that v is a representative vector for the point $[v] \in P(V)$. Clearly if $\lambda \neq 0$ then λv is another representative vector so

$$[\lambda v] = [v].$$

Now suppose we choose a basis $\{v_0, \ldots, v_n\}$ for V. The vector v can be written

$$v = \sum_{i=0}^{n} x_i v_i$$

and the $n+1$-tuple $(x_0, \ldots, x_n)$ provides the coordinates of $v \in V$. If $v \neq 0$ we write the corresponding point $[v] \in P(V)$ as $[v] = [x_0, x_1, \ldots, x_n]$ and these are known as homogeneous coordinates for a point in $P(V)$. Again, for $\lambda \neq 0$

$$[\lambda x_0, \lambda x_1, \ldots, \lambda x_n] = [x_0, x_1, \ldots, x_n].$$

Homogeneous coordinates give us another point of view of projective space. Let $U_0 \subset P(V)$ be the subset of points with homogeneous coordinates $[x_0, x_1, \ldots, x_n]$ such that $x_0 \neq 0$. (Since if $\lambda \neq 0$, $x_0 \neq 0$ if and only if $\lambda x_0 \neq 0$, so this is a well-defined subset, independent of the choice of $(x_0, \ldots, x_n)$). Then, in U_0,

$$[x_0, x_1, \ldots, x_n] = [x_0, x_0(x_1/x_0), \ldots, x_0(x_n/x_0)] = [1, (x_1/x_0), \ldots, (x_n/x_0)].$$

Thus we can uniquely represent any point in U_0 by one of the form $[1, y_1, \ldots, y_n]$, so

$$U_0 \cong \mathbf{C}^n.$$

The points we have missed out are those for which $x_0 = 0$, but these are the 1-dimensional subspaces of the n-dimensional vector subspace spanned by $v_1, \ldots, v_n$, which is a projective space of one lower dimension. So we can write

$$\mathbf{P}_n = \mathbf{C}^n \cup \mathbf{P}_{n-1}$$

A large chunk of complex projective n-space is thus our familiar $\mathbf{C}^n$.

Example: The simplest example of this is the case $n = 1$. Since a one-dimensional projective space is a single point (if $\dim V = 1$, V is the only 1-dimensional subspace) the projective line $\mathbf{P}_1 = \mathbf{C} \cup pt$. Since $[x_0, x_1]$ maps to $x_1/x_0 \in \mathbf{C}$ we usually call this extra point $[0, 1]$ the point ∞. The projective line is what is called in complex analysis the *extended complex plane* $\mathbf{C} \cup \{\infty\}$.

Having said that, there are many different copies of $\mathbf{C}^n$ inside $\mathbf{P}_n$, for we could have chosen x_i instead of x_0, or coordinates with respect to a totally different basis. Projective space should normally be thought of as a homogeneous object, without any distinguished copy of $\mathbf{C}^n$ inside.

We defined U_0 above as the subset of $\mathbf{P}_n$ where $x_0 \neq 0$, so we can similarly define subsets U_i where $x_i \neq 0$. Any point of $\mathbf{P}_n$ lies in one of these sets. We can use them to make $\mathbf{P}_n$ into a topological space – define a set V to be open if $V \cap U_i$ is an open set in $\mathbf{C}^n$ under the map

$$[x_0, x_1, \ldots, x_n] \mapsto (x_0/x_i, \ldots, 1, \ldots x_n/x_i).$$

1.2 Linear subspaces

Definition 3 *A linear subspace of the projective space $P(V)$ is the set of 1-dimensional vector subspaces of a vector subspace $U \subseteq V$.*

Note that a linear subspace is a projective space in its own right, the projective space $P(U)$.

Recall that a 1-dimensional projective space is called a projective line. We have the following two propositions which show that projective lines behave nicely:

Proposition 1 *Through any two distinct points in a projective space there passes a unique projective line.*

Proof: Let $P(V)$ be the projective space and $x, y \in P(V)$ distinct points. Let u, v be representative vectors. Then u, v are linearly independent for otherwise $u = \lambda v$ and

$$x = [u] = [\lambda v] = [v] = y.$$

Let $U \subseteq V$ be the 2-dimensional vector space spanned by u and v, then $P(U) \subset P(V)$ is a line containing x and y.

Suppose $P(U')$ is another such line, then $u \in U'$ and $v \in U'$ and so the space spanned by u, v (namely U) is a subspace of U'. But U and U' are 2-dimensional so $U = U'$ and the line is thus unique. □

Proposition 2 *In a projective plane, two distinct projective lines intersect in a unique point.*

Proof: Let the projective plane be $P(V)$ where $\dim V = 3$. Two lines are defined by $P(U_1), P(U_2)$ where U_1, U_2 are distinct 2-dimensional subspaces of V. Now from elementary linear algebra

$$\dim V \geq \dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

so that

$$3 \geq 2 + 2 - \dim(U_1 \cap U_2)$$

and

$$\dim(U_1 \cap U_2) \geq 1.$$

But since U_1 and U_2 are 2-dimensional,

$$\dim(U_1 \cap U_2) \leq 2$$

with equality if and only if $U_1 = U_2$. As the lines are distinct, equality doesn't occur and so we have the 1-dimensional vector subspace

$$U_1 \cap U_2 \subset V$$

which is the required point of intersection in $P(V)$. □

Any 2-dimensional subspace of $\mathbf{C}^3$ is defined by a single equation

$$a_0 x_0 + a_1 x_1 + a_2 x_2 = 0$$

and if a_1 and a_2 are not both zero, this intersects $U_0 \cong \mathbf{C}^2$ (the points where $x_0 \neq 0$) where

$$0 = a_0 + a_1(x_1/x_0) + a_2(x_2/x_0) = a_0 + a_1 y_1 + a_2 y_2$$

which is an ordinary line in $\mathbf{C}^2$ with coordinates y_1, y_2. The projective line has one extra point on it, where $x_0 = 0$, i.e. the point $[0, a_2, -a_1]$. Conversely, any line in $\mathbf{C}^2$ extends uniquely to a projective line in $\mathbf{P}_2$.

Two lines in $\mathbf{C}^2$ are parallel if they are of the form

$$a_0 + a_1 y_1 + a_2 y_2 = 0, \qquad b_0 + a_1 y_1 + a_2 y_2 = 0$$

but then the added point to make them projective lines is the same one: $[0, a_2, -a_1]$, so the two lines meet at a single point on the "line at infinity" $\mathbf{P}_1$.

One of our goals will be to have good theorems about the number of points of intersection of two curves, so introducing the projective plane gives us a start: *any* two distinct lines meet in a single point.

1.3 Projective transformations

If V, W are vector spaces and $T : V \to W$ is a linear transformation, then a vector subspace $U \subseteq V$ gets mapped to a vector subspace $T(U) \subseteq W$. If T has a non-zero kernel, $T(U)$ may have dimension less than that of U, but if $\ker T = 0$ then $\dim T(U) = \dim U$. In particular, if U is one-dimensional, so is $T(U)$ and so T gives a well-defined map

$$\tau : P(V) \to P(W).$$

Definition 4 *A projective transformation from $P(V)$ to $P(W)$ is the map τ defined by an invertible linear transformation $T : V \to W$.*

Note that if $\lambda \neq 0$, then λT and T define the same linear transformation since

$$[(\lambda T)(v)] = [\lambda(T(v))] = [T(v)].$$

The converse is also true: suppose T and T' define the same projective transformation τ. Take a basis $\{v_0, \ldots, v_n\}$ for V, then since

$$\tau([v_i]) = [T'(v_i)] = [T(v_i)]$$

we have

$$T'(v_i) = \lambda_i T(v_i)$$

for some non-zero scalars λ_i and also

$$T'(\sum_{i=0}^{n} v_i) = \lambda T(\sum_{i=0}^{n} v_i)$$

for some non-zero λ. But then

$$\sum_{i=0}^{n} \lambda T(v_i) = \lambda T(\sum_{i=0}^{n} v_i) = T'(\sum_{i=0}^{n} v_i) = \sum_{i=0}^{n} \lambda_i T(v_i).$$

Since T is invertible, $T(v_i)$ are linearly independent, so this implies $\lambda_i = \lambda$. Then $T'(v_i) = \lambda T(v_i)$ for all basis vectors and hence for all vectors and so

$$T' = \lambda T.$$

Example: You are, in fact, already familiar with one class of projective transformations – Möbius transformations of the extended complex plane. These are just projective transformations of the complex projective line $\mathbf{P}_1$ to itself. We describe points in $\mathbf{P}_1$ by homogeneous coordinates $[z_0, z_1]$, and then a projective transformation τ is given by

$$\tau([z_0, z_1]) = ([az_0 + bz_1, cz_0 + dz_1])$$

where $ad - bc \neq 0$. This corresponds to the invertible linear transformation

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It is convenient to write $\mathbf{P}_1 = \mathbf{C} \cup \{\infty\}$ where the point ∞ is now the 1-dimensional space $z_1 = 0$. Then if $z_1 \neq 0$, $[z_0, z_1] = [z, 1]$ and

$$\tau([z, 1]) = [az + b, cz + d]$$

and if $cz + d \neq 0$ we can write

$$\tau([z, 1]) = \left[\frac{az + b}{cz + d}, 1\right]$$

which is the usual form of a Möbius transformation, i.e.

$$z \mapsto \frac{az + b}{cz + d}.$$

The advantage of projective geometry is that the point $\infty = [1, 0]$ plays no special role. If $cz + d = 0$ we can still write

$$\tau([z, 1]) = [az + b, cz + d] = [az + b, 0] = [1, 0]$$

and if $z = \infty$ (i.e. $[z_0, z_1] = [1, 0]$) then we have

$$\tau([1, 0]) = [a, c].$$

Example: If we view the projective plane $\mathbf{P}_2$ in the same way, we get some less familiar transformations. Write

$$\mathbf{P}_2 = \mathbf{C}^2 \cup \mathbf{P}_1$$

where the projective line at infinity is $x_0 = 0$. A linear transformation $T : \mathbf{C}^3 \to \mathbf{C}^3$ can then be written as the matrix

$$T = \begin{pmatrix} d & b_1 & b_2 \\ c_1 & a_{11} & a_{12} \\ c_2 & a_{21} & a_{22} \end{pmatrix}$$

and its action on $[1, x, y]$ can be expressed, with $\mathbf{v} = (x, y) \in \mathbf{C}^2$, as

$$\mathbf{v} \mapsto \frac{1}{\mathbf{b} \cdot \mathbf{v} + d} (A\mathbf{v} + \mathbf{c})$$

where A is the 2×2 matrix a_{ij} and $\mathbf{b}, \mathbf{c}$ the vectors $(b_1, b_2), (c_2, c_2)$. These are the 2-dimensional versions of Möbius transformations. Each one can be considered as a composition of

- an invertible linear transformation $\mathbf{v} \mapsto A\mathbf{v}$
- a translation $\mathbf{v} \mapsto \mathbf{v} + \mathbf{c}$
- an inversion $\mathbf{v} \mapsto \mathbf{v}/(\mathbf{b} \cdot \mathbf{v} + d)$

Clearly it is easier here to consider projective transformations defined by 3×3 matrices, just ordinary linear algebra.

However, we can see from this viewpoint that the projective transformations that take $\mathbf{C}^2$ to itself are those with $b_1 = b_2 = 0$ and then the action on $\mathbf{C}^2$ is

$$\mathbf{v} \mapsto d^{-1}(A\mathbf{v} + c)$$

– a combination of linear transformation and translation. These are known as *affine transformations*, so in a natural way the complement of a line in $\mathbf{P}_2$ is an affine space (a "vector space without a distinguished origin".) The subsets U_i are called affine subsets of $\mathbf{P}_2$.

Example: A more geometric example of a projective transformation is to take two lines $P(U), P(U')$ in a projective plane $P(V)$ and let $O \in P(V)$ be a point disjoint from both. For each point $x \in P(U)$, the unique line joining O to x intersects $P(U')$ in a unique point $X = \tau(x)$. Then

$$\tau : P(U) \to P(U')$$

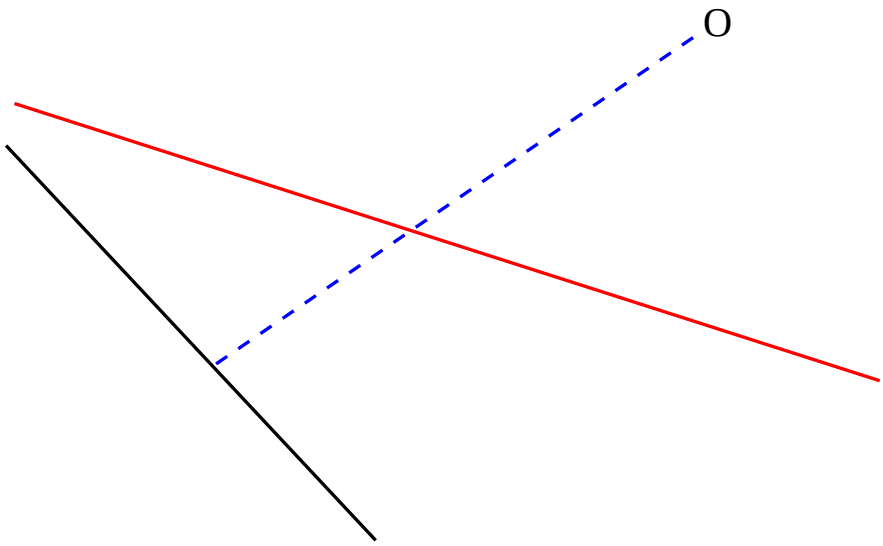
is a projective transformation.

To see why, let W be the 1-dimensional subspace of V defined by $O \in P(V)$. Then since O does not lie in $P(U')$, $W \cap U' = 0$. This means that

$$V = W \oplus U'.$$

Now take $a \in U$ as a representative vector for x. It can be expressed uniquely as $a = w + a'$, with $w \in W$ and $a' \in U'$. The projective line joining O to x is defined by the 2-dimensional vector subspace of V spanned by w and a and so $a' = a - w$ is a representative vector for $\tau(x)$. In linear algebra terms the map $a \mapsto a'$ is just the linear projection map $P : V \to U'$ restricted to U. It has zero kernel since O does not lie in $P(U)$, and hence $W \cap U = 0$. Thus $T : U \to U'$ is an isomorphism and τ is a projective transformation.

If we restrict to the points in $\mathbf{R}^2$, then this is what this *projection from* O looks like:



A linear transformation of a vector space of dimension n is determined by its value on n linearly independent vectors. A similar statement holds in projective space. The analogue of linear independence is the following

Definition 5 *Let $P(V)$ be an n-dimensional projective space, then $n+2$ points in $P(V)$ are said to be in general position if each subset of $n+1$ points has representative vectors in V which are linearly independent.*

Example: Any two distinct points in a projective line are represented by linearly independent vectors, so any three distinct points are in general position.

Theorem 3 *If $X_1, \ldots, X_{n+2}$ are in general position in $P(V)$ and $Y_1, \ldots, Y_{n+2}$ are in general position in $P(W)$, then there is a unique projective transformation $\tau : P(V) \to P(W)$ such that $\tau(X_i) = Y_i$, $1 \leq i \leq n+2$.*

Proof: First choose representative vectors $v_1, \ldots, v_{n+2} \in V$ for the points $X_1, \ldots, X_{n+2}$ in $P(V)$. By general position the first $n+1$ vectors are linearly independent, so they form a basis for V and there are scalars λ_i such that

$$v_{n+2} = \sum_{i=1}^{n+1} \lambda_i v_i \tag{1}$$

If $\lambda_i = 0$ for some i, then (1) provides a linear relation amongst a subset of $n+1$ vectors, which is not possible by the definition of general position, so we deduce that

$\lambda_i \neq 0$ for all i. This means that each $\lambda_i v_i$ is also a representative vector for X_i, so (1) tells us that we could have chosen representative vectors v_i such that

$$v_{n+2} = \sum_{i=1}^{n+1} v_i \tag{2}$$

Moreover, given v_{n+2}, these v_i are unique for

$$\sum_{i=1}^{n+1} v_i = \sum_{i=1}^{n+1} \mu_i v_i$$

implies $\mu_i = 1$ since $v_1, \ldots, v_{n+1}$ are linearly independent.

[*Note: This is a very useful idea which can simplify the solution of many problems*].

Now do the same for the points $Y_1, \ldots Y_{n+2}$ in $P(W)$ and choose representative vectors such that

$$w_{n+2} = \sum_{i=1}^{n+1} w_i \tag{3}$$

Since $v_1, \ldots, v_{n+1}$ are linearly independent, they form a basis for V so there is a unique linear transformation $T : V \to W$ such that $Tv_i = w_i$ for $1 \leq i \leq n+1$. Since $w_1, \ldots, w_{n+1}$ are linearly independent, T is invertible. Furthermore, from (2) and (3)

$$Tv_{n+2} = \sum_{i=1}^{n+1} Tv_i = \sum_{i=1}^{n+1} w_i = w_{n+2}$$

and so T defines a projective transformation τ such that $\tau(X_i) = Y_i$ for all $n+2$ vectors v_i.

To show uniqueness, suppose T' defines another projective transformation τ' with the same property. Then $T'v_i = \mu_i w_i$ and

$$\mu_{n+2} w_{n+2} = T'v_{n+2} = \sum_{i=1}^{n+1} T'v_i = \sum_{i=1}^{n+1} \mu_i w_i.$$

But by the uniqueness of the representation (3), we must have $\mu_i/\mu_{n+2} = 1$, so that $T'v_i = \mu_{n+2} T v_i$ and $\tau' = \tau$. □

Examples:

1. In $\mathbf{P}_1$ take the three distinct points $[0,1],[1,1],[1,0]$ and any other three distinct points X_1, X_2, X_3. Then there is a unique projective transformation taking X_1, X_2, X_3 to $[0,1],[1,1],[1,0]$. In the language of complex analysis, we can say that there is a unique Möbius transformation taking any three distinct points to $0, 1, \infty$.

2. In any projective line we could take the three points $[0,1],[1,1],[1,0]$ and then for X_1, X_2, X_3 any permutation of these. Now projective transformations of a space to itself form a group under composition, so we see that the group of projective transformations of a line to itself always contains a copy of the symmetric group S_3. In fact if we take the scalars to be the field $\mathbf{Z}_2$ with two elements 0 and 1, the *only* points on the projective line are $[0,1],[1,1],[1,0]$, and S_3 is the full group of projective transformations.

As an example of the use of the notion of general position, here is a classical theorem called Desargues' theorem. In fact, Desargues (1591-1661) is generally regarded as the founder of projective geometry. The proof we give here uses the method of choosing representative vectors above.

Theorem 4 *(Desargues) Let A, B, C, A', B', C' be distinct points in a projective space $P(V)$ such that the lines AA', BB' CC' are distinct and concurrent. Then the three points of intersection $AB \cap A'B'$, $BC \cap B'C'$, $CA \cap C'A'$ are collinear.*

Proof: Let P be the common point of intersection of the three lines AA', BB', CC'. Since P, A, A' lie on a projective line and are distinct, they are in general position, so as in (2) we choose representative vectors p, a, a' such that

$$p = a + a'.$$

These are vectors in a 2-dimensional subspace of V. Similarly we have representative vectors b, b' for B, B' and c, c' for C, C' with

$$p = b + b' \qquad p = c + c'.$$

It follows that $a + a' = b + b'$ and so

$$a - b = b' - a' = c''$$

and similarly

$$b - c = c' - b' = a'' \qquad c - a = a' - c' = b''.$$

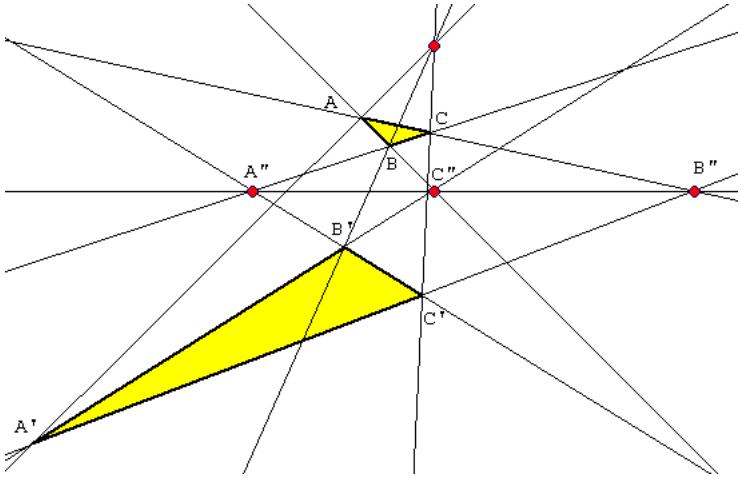
But then

$$c'' + a'' + b'' = a - b + b - c + c - a = 0$$

and a'', b'', c'' are linearly dependent and lie in a 2-dimensional subspace of V. Hence the points A'', B'', C'' in $P(V)$ represented by a'', b'', c'' are collinear.

Now since $c'' = a - b$, c'' lies in the 2-dimensional space spanned by a and b, so C'' lies on the line AB. Since c'' also equals $b' - a'$, C'' lies on the line $A'B'$ and so c'' represents the point $AB \cap A'B'$. Repeating for B'' and A'' we see that these are the three required collinear points. □

Desargues' theorem is a theorem in projective space which we just proved by linear algebra – linear independence of vectors. However, if we take the projective space $P(V)$ to be the real projective plane $P^2(\mathbf{R})$ and then just look at that part of the data which lives in $\mathbf{R}^2$, we get a theorem about perspective triangles in the plane:



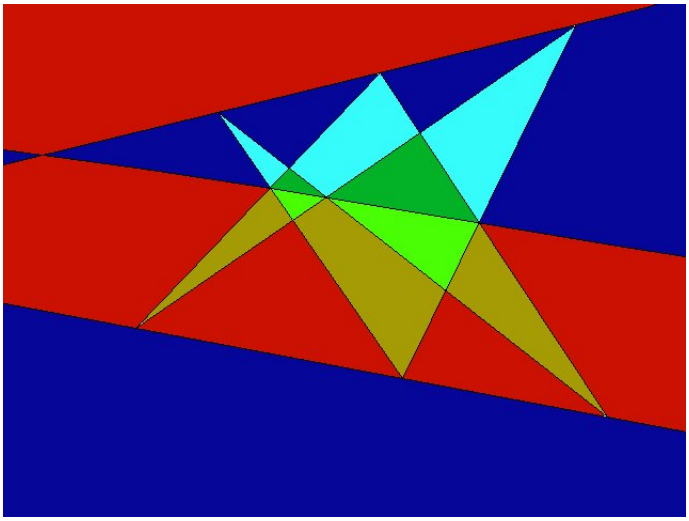
Here is an example of the use of projective geometry – a "higher form of geometry" to prove simply a theorem in $\mathbf{R}^2$ which is less accessible by other means. Another theorem in the plane for which these methods give a simple proof is Pappus' theorem. Pappus of Alexandria (290-350) was thinking again of plane Euclidean geometry, but his theorem makes sense in the projective plane since it only discusses collinearity and not questions about angles and lengths. It means that we can transform the given configuration by a projective transformation to a form which reduces the proof to simple linear algebra calculation:

Theorem 5 *(Pappus) Let A, B, C and A', B', C' be two pairs of collinear triples of distinct points in a projective plane. Then the three points $BC' \cap B'C, CA' \cap C'A, AB' \cap A'B$ are collinear.*

Proof: Without loss of generality, we can assume that A, B, C', B' are in general position. If not, then two of the three required points coincide, so the conclusion is trivial. By Theorem 3, we can then assume that

$$A = [1, 0, 0], \quad B = [0, 1, 0], \quad C' = [0, 0, 1], \quad B' = [1, 1, 1].$$

The line AB is defined by the 2-dimensional subspace $\{(x_0, x_1, x_2) \in F^3 : x_2 = 0\}$, so the point C, which lies on this line, is of the form $C = [1, c, 0]$ and $c \neq 0$ since $A \neq C$. Similarly the line $B'C'$ is $x_0 = x_1$, so $A' = [1, 1, a]$ with $a \neq 1$.



The line BC' is defined by $x_0 = 0$ and $B'C$ is defined by the span of $(1, 1, 1)$ and $(1, c, 0)$, so the point $BC' \cap B'C$ is represented by the linear combination of $(1, 1, 1)$ and $(1, c, 0)$ for which $x_0 = 0$, i.e.

$$(1, 1, 1) - (1, c, 0) = (0, 1 - c, 1).$$

The line $C'A$ is given by $x_1 = 0$, so similarly $CA' \cap C'A$ is represented by

$$(1, c, 0) - c(1, 1, a) = (1 - c, 0, -ca).$$

Finally AB' is given by $x_1 = x_2$, so $AB' \cap A'B$ is

$$(1, 1, a) + (a - 1)(0, 1, 0) = (1, a, a).$$

But then

$$(c - 1)(1, a, a) + (1 - c, 0, -ca) + a(0, 1 - c, 1) = 0.$$

Thus the three vectors span a 2-dimensional subspace and so the three points lie on a projective line. □

2 Plane curves

2.1 Basic definitions

From now on we will mainly work in the projective plane $\mathbf{P}_2$ and use homogeneous coordinates $[x, y, z]$ instead of $[x_0, x_1, x_2]$. The equation of a line is

$$ax + by + cz = 0$$

and this gives a well-defined subset of $\mathbf{P}_2$ because if we replace (x, y, z) by $(\lambda x, \lambda y, \lambda z)$ the equation still holds.

Definition 6 *A polynomial $P(x, y, z)$ is homogeneous of degree d if*

$$P(\lambda x, \lambda y, \lambda z) = \lambda^d P(x, y, z).$$

Clearly $P(x, y, z) = 0$ is a well-defined subset of $\mathbf{P}_2$.

Definition 7 *Let $P(x, y, z)$ be a homogeneous polynomial of degree $d > 0$ with no repeated factors, then $P(x, y, z) = 0$ defines a plane projective curve C in $\mathbf{P}_2$.*

Remark: 1. The subset $P(x, y, z) = 0$ in $\mathbf{P}_2$ is non-empty because fixing y and z we have a polynomial in x which over $\mathbf{C}$ always has roots. This is why we work over the complex numbers – $P(x, y, z) = x^2 + y^2 + z^2$ defines an algebraic curve over $\mathbf{R}$ but it has no real points.

2. The condition of having no repeated factors is to ensure that the polynomial is uniquely determined (up to a scalar multiple) by the curve C – obviously $P(x, y, z)$ and $P^2(x, y, z)$ define the same subset. The fact that with this condition the zeros determine P follows from Hilbert's Nullstellensatz, which is outside the scope of this course.

Definition 8 *The curve $C \subset \mathbf{P}_2$ is said to be irreducible if P has no non-constant factors other than a scalar multiple of itself. An irreducible plane curve D is said to be a component of C if its defining polynomial Q divides P.*

Remark: Note that this use of the word "component" is different from the topological one – two lines given by linear polynomials P, Q have non-empty intersection so the reducible curve C defined by $PQ = 0$ is a connected topological space.

Definition 9 *The point* $[a,b,c] \in \mathbf{P}_2$ *is a singular point of* C *if* $P(a,b,c) = 0$ *and*

$$\frac{\partial P}{\partial x}(a,b,c) = \frac{\partial P}{\partial y}(a,b,c) = \frac{\partial P}{\partial z}(a,b,c) = 0.$$

Remark: For a homogeneous function P, differentiating $P(\lambda x, \lambda y, \lambda z) = \lambda^d P(x,y,z)$ with respect to λ leads via the chain rule directly to Euler's relation:

$$x\frac{\partial P}{\partial x} + y\frac{\partial P}{\partial y} + z\frac{\partial P}{\partial z} = dP$$

so that the vanishing of the partial derivatives actually implies the vanishing of P.

Example: 1. The reducible curve C defined by a pair of lines has a unique singular point – the point of intersection – since

$$\frac{\partial PQ}{\partial x} = P\frac{\partial Q}{\partial x} + Q\frac{\partial P}{\partial x}$$

and this, and the other partial derivatives, vanish when $P = Q = 0$.

2. The curve $x^2 + y^2 + z^2 = 0$ has no singularities since the vanishing of the partial derivatives gives $x = y = z = 0$ but this does not define a point in projective space.

If $[a,b,c]$ is not a singular point of C then at least one partial derivative is non-vanishing. Then we define:

Definition 10 *If* $p = [a,b,c]$ *is a nonsingular point of* C*, the tangent line at* p *is defined by the equation*

$$x\frac{\partial P}{\partial x}(a,b,c) + y\frac{\partial P}{\partial y}(a,b,c) + z\frac{\partial P}{\partial z}(a,b,c) = 0.$$

At this point it may be useful to view projective curves in terms of their intersections with one of the affine open sets U_i, say U_2 where $z \neq 0$. Then identifying U_2 with $\mathbf{C}^2$ via $(x/z, y/z)$ the curve $C \cap U_2$ is defined by the nonhomogeneous equation

$$P(x,y,1) = 0.$$

Conversely, the subset of $\mathbf{C}^2$ defined by a polynomial $p(x,y) = 0$ can be extended to a homogeneous polynomial by

$$P(x,y,z) = z^d p(\frac{x}{z}, \frac{y}{z})$$

so long as $d \geq m + n$ where $x^m y^n$ is the highest power in p. If the line $z = 0$ is not a component of C then this gives a one-to-one correspondence, but the extra points on $C \cap \{z = 0\}$ have to be analyzed by using one of the other open sets U_0, U_1.

Our definition of tangent line makes more sense in this affine open set, where we might define the "normal direction" to $p(x, y) = 0$ at (a, b) by $(\partial p/\partial x, \partial p/\partial y))$ and the tangent would then be the line

$$(x - a)\frac{\partial p}{\partial x}(a, b) + (y - b)\frac{\partial p}{\partial y}(a, b) = 0.$$

It is easy to check that this line is the intersection with U_2 of the tangent line according to Definition 10.

Example: The parabola $y^2 = 4x$ in $\mathbf{R}^2$ extends to the curve C of degree 2 in $P(\mathbf{R}^3)$ defined by the homogeneous polynomial

$$P(x, y, z) = y^2 - 4xz.$$

We can rewrite this as

$$y^2 - (x + z)^2 + (x - z)^2$$

so its intersection with U_1, where $y \neq 0$ is the curve

$$(x + z)^2 - (x - z)^2 = 1$$

which in coordinates $(x+z, x-z)$ is a hyperbola. From the point of view of projective geometry the curve C intersects the line $z = 0$ in the single real point $[1, 0, 0]$, and the line $y = 0$ in two points $[1, 0, 0], [0, 0, 1]$. In the affine set $x + z \neq 0$, we have coordinates $(y/(x + z), (x - z)/(x + z))$ and then

$$\frac{y^2}{(x + z)^2} + \frac{(x - z)^2}{(x + z)^2} = 1$$

which is the equation of a circle. So in projective space all these curves become one and the same type.

2.2 Conics

A conic is a plane projective curve of degree 2. It is defined by a homogeneous quadratic form $P(x, y, z)$. This is algebraically the same thing as:

Definition 11 *A symmetric bilinear form on a vector space V is a map $B : V \times V \to F$ such that*

- $B(v, w) = B(w, v)$
- $B(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 B(v_1, w) + \lambda_2 B(v_2, w)$

The form is said to be nondegenerate if $B(v, w) = 0$ for all $w \in V$ implies $v = 0$.

If we take a basis $v_1, \ldots, v_n$ of V, then $v = \sum_i x_i v_i$ and $w = \sum_i y_i v_i$ so that

$$B(v, w) = \sum_{i,j} B(v_i, v_j) x_i y_j$$

and so is uniquely determined by the symmetric matrix $\beta_{ij} = B(v_i, v_j)$. The bilinear form is nondegenerate if and only if β_{ij} is nonsingular.

We can add symmetric bilinear forms: $(B+C)(v, w) = B(v, w)+C(v, w)$ and multiply by a scalar $(\lambda B)(v, w) = \lambda B(v, w)$ so they form a vector space isomorphic to the space of symmetric $n \times n$ matrices which has dimension $n(n + 1)/2$. If we take a different basis

$$w_i = \sum_j P_{ji} v_j$$

then

$$B(w_i, w_j) = B(\sum_k P_{ki} v_k, \sum_\ell P_{\ell j} v_\ell) = \sum_{k,\ell} P_{ki} B(v_k, v_\ell) P_{\ell j}$$

so that the matrix $\beta_{ij} = B(v_i, v_j)$ changes under a change of basis to

$$\beta' = P^T \beta P.$$

Over the real or complex numbers we can divide by 2 and then we often speak of the quadratic form $B(v, v)$ which determines the bilinear form since

$$B(v + w, v + w) = B(v, v) + B(w, w) + 2B(v, w)$$

Here we have the basic result:

Theorem 6 *Let B be a quadratic form on a complex vector space V of dimension n. Then there is a basis such that if $v = \sum_i z_i v_i$*

$$B(v, v) = \sum_{i=1}^{m} z_i^2.$$

If B is non-degenerate then $m = n$.

Proof: The proof is elementary – just *completing the square*. We note that changing the basis is equivalent to changing the coefficients x_i of v by an invertible linear transformation.

First we write down the form in one basis, so that

$$B(v,v) = \sum_{i,j} \beta_{ij} x_i x_j$$

and ask: *is there a term* $\beta_{ii} \neq 0$*?*. If not, then we create one. If the coefficient of $x_i x_j$ is non-zero, then putting $y_i = (x_i + x_j)/2, y_j = (x_i - x_j)/2$ we have

$$x_i x_j = y_i^2 - y_j^2$$

and so we get a term $\beta'_{ii} \neq 0$.

If there is a term $\beta_{ii} \neq 0$, then we note that

$$\frac{1}{\beta_{ii}}(\beta_{i1}x_1 + \ldots + \beta_{in}x_n)^2 = \beta_{ii}x_i^2 + 2\sum_{k \neq i} \beta_{ik} x_k x_i + R$$

where R involves the x_k with $k \neq i$. So if

$$y_i = \beta_{i1}x_1 + \ldots + \beta_{in}x_n$$

then

$$B(v,v) = \frac{1}{\beta_{ii}} y_i^2 + B_1$$

where B_1 is a quadratic form in the $n-1$ variables x_k, $k \neq i$.

We now repeat the procedure to find a basis such that if v has coefficients $y_1, \ldots, y_n$, then

$$B(v,v) = \sum_{i=1}^{m} c_i y_i^2.$$

Over $\mathbf{C}$ we can write $z_i = \sqrt{c_i} y_i$ and get a sum of squares. □

Example: Consider the quadratic form in $\mathbf{C}^3$

$$B(v,v) = xy + yz + zx.$$

We put

$$y_1 = (x+y)/2, \qquad y_2 = (x-y)/2$$

to get

$$B(v,v) = y_1^2 - y_2^2 + z(2y_1).$$

Now complete the square:

$$B(v,v) = (y_1 + z)^2 - y_2^2 - z^2$$

so that with $z_1 = y_1 + z, z_2 = iy_2, z_3 = iz$ we have $z_1^2 + z_2^2 + z_3^2$.

An invertible linear transformation of $\mathbf{C}^3$ defines a projective transformation of $\mathbf{P}_2$ so Theorem 6 tells us that any conic is equivalent to one of the following by a projective transformation:

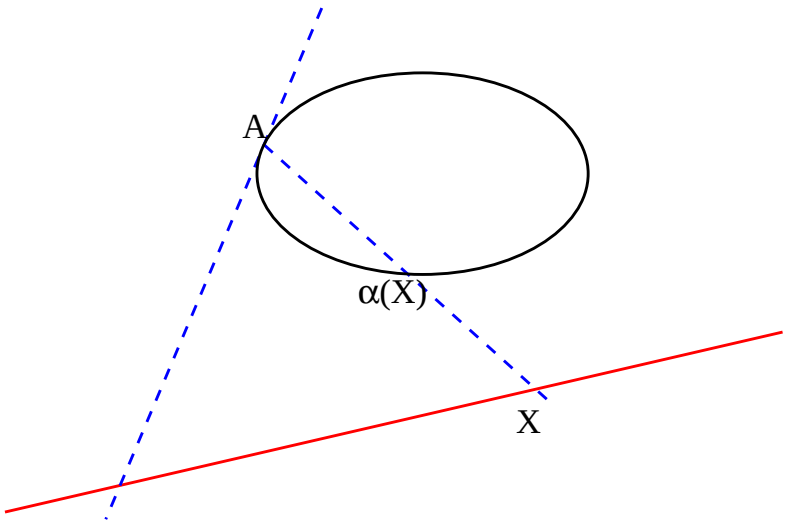
- $P(x,y,z) = x^2$ – this is the (double) line $x = 0$
- $P(x,y,z) = x^2 + y^2$ – this is a pair of lines $x + iy = 0, x - iy = 0$
- $P(x,y,z) = x^2 + y^2 + z^2$ – a nonsingular conic.

2.3 Rational parametrization of the conic

Theorem 7 *Let C be a nonsingular conic in a projective plane $P(V)$ over the field F, and let A be a point on C. Let $P(U) \subset P(V)$ be a projective line not containing A. Then there is a bijection*

$$\alpha : P(U) \to C$$

such that, for $X \in P(U)$,the points $A, X, \alpha(X)$ are collinear.



Proof: Suppose the conic is defined by the nondegenerate symmetric bilinear form B. Let $a \in V$ be a representative vector for A, then $B(a,a) = 0$ since A lies on the conic. Let $x \in P(U)$ be a representative vector for $X \in P(U)$. Then a and x are linearly independent since X does not lie on the line $P(U)$. Extend a, x to a basis a, x, y of V.

Now B restricted to the space spanned by a, x is not identically zero, because if it were, the matrix of B with respect to this basis would be of the form

$$\begin{pmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{pmatrix}$$

which is singular. So at least one of $B(x,x)$ and $B(a,x)$ is non-zero.

Any point on the line AX is represented by a vector of the form $\lambda a + \mu x$ and this lies on the conic C if

$$0 = B(\lambda a + \mu x, \lambda a + \mu x) = 2\lambda\mu B(a,x) + \mu^2 B(x,x).$$

When $\mu = 0$ we get the point X. The other solution is $2\lambda B(a,x) + \mu B(x,x) = 0$ i.e. the point with representative vector

$$w = B(x,x)a - 2B(a,x)x \tag{4}$$

which is non-zero since the coefficients are not both zero.

We define the map $\alpha : P(U) \to C$ by

$$\alpha(X) = [w]$$

which has the collinearity property of the statement of the Theorem. If $Y \in C$ is distinct from A, then the line AY meets the line $P(U)$ in a unique point, so α^{-1} is well-defined on this subset. By the definition of α in (4), $\alpha(X) = A$ if and only if $B(a,x) = 0$. Since B is nonsingular $f(x) = B(a,x)$ is a non-zero linear map from V to F and so defines a line (the tangent to C at A), which hence meets $P(U)$ in one point. Thus α has a well-defined inverse and is therefore a bijection. □

Example: Consider the case of the conic

$$x^2 + y^2 - z^2 = 0.$$

Take $A = [1,0,1]$ and the line $P(U)$ defined by $x = 0$. Note that this conic and the point and line are defined over any field since the coefficients are 0 or 1.

A point $X \in P(U)$ is of the form $X = [0, 1, t]$ or $[0, 0, 1]$ and the map α is

$$\begin{aligned} \alpha([0,1,t]) &= [B((0,1,t),(0,1,t))(1,0,1) - 2B((1,0,1),(0,1,t))(0,1,t)] \\ &= [1-t^2, 2t, 1+t^2] \end{aligned}$$

or $\alpha([0, 0, 1]) = [-1, 0, 1]$.

This has an interesting application if we use the field of rational numbers $F = \mathbf{Q}$. Suppose we want to find all right-angled triangles whose sides are of integer length. By Pythagoras, we want to find positive integer solutions to

$$x^2 + y^2 = z^2.$$

But then $[x, y, z]$ is a point on the conic. Conversely, if $[x_0, x_1, x_2]$ lies on the conic, then multiplying by the least common multiple of the denominators of the rational numbers x_0, x_1, x_2 gives integers such that $[x, y, z]$ is on the conic.

But what we have seen is that *any* point on the conic is either $[-1, 0, 1]$ or of the form

$$[x, y, z] = [1 - t^2, 2t, 1 + t^2]$$

for some rational number $t = p/q$, so we get all integer solutions by putting

$$x = q^2 - p^2, \quad y = 2pq, \quad z = q^2 + p^2.$$

For example, $p = 1, q = 2$ gives $3^2 + 4^2 = 5^2$ and $p = 2, q = 3$ gives $5^2 + 12^2 = 13^2$.

One other consequence of Theorem 7 is that we can express a point (x, y) on the general conic

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

in the form

$$x = \frac{p(t)}{r(t)}, \qquad y = \frac{q(t)}{r(t)}$$

where p, q and r are quadratic polynomials in t. Writing x, y as rational functions of t is why the process we have described is sometimes called the *rational parametrization of the conic*. It has its uses in integration. We can see, for example, that

$$\int \frac{dx}{x + \sqrt{ax^2 + bx + c}}$$

can be solved by elementary functions because if $y = x + \sqrt{ax^2 + bx + c}$ then

$$(y - x)^2 - ax^2 - bx - c = 0$$

and this is the equation of a conic. We can solve it by $x = p(t)/r(t), y = q(t)/r(t)$ and with this substitution, the integral becomes

$$\int \frac{r'(t)p(t) - p'(t)r(t)}{q(t)r(t)} dt$$

and expanding the rational integrand into partial fractions we get rational and logarithmic terms after integration.

3 Intersections of curves

3.1 Resultants

We can gain quite a lot of information about curves from Bézout's Theorem which says roughly that curves C and D of degrees n and m respectively intersect in mn points. This is clearly not true in general (for example a line which is tangent to a nonsingular conic meets it in one point, not two). This means that we have to do some work to define carefully the intersection multiplicity and the conditions under which the theorem holds.

The basic idea is to write

$$\begin{aligned} P(x,y,z) &= a_0(y,z) + a_1(y,z)x + \ldots + a_n(y,z)x^n \\ Q(x,y,z) &= b_0(y,z) + b_1(y,z)x + \ldots + b_m(y,z)x^m \end{aligned}$$

and forget for the moment the dependence of the coefficients on y, z. The condition that two polynomials in x have a common root gives a polynomial relation on the a_i, b_j and, putting y and z back in, we can find the number of solutions.

Let

$$\begin{aligned} p(x) &= a_0 + a_1x + \ldots + a_nx^n \\ q(x) &= b_0 + b_1x + \ldots + b_mx^m \end{aligned}$$

be two complex polynomials in x with $a_n \neq 0$. If they have a common factor ℓ then $p = \ell r, q = \ell s$ for polynomials r, s of degree $\leq n-1, m-1$ respectively, and so

$$sp - rq = 0.$$

This is a linear equation for the n coefficients of r and m coefficients of s. But a polynomial of the form $sp - rq$ has degree $\leq m+n-1$ and so has $m+n$ coefficients, hence $sp - rq = 0$ is an equation of the form $Av = 0$ for a vector $v \in \mathbf{C}^{m+n}$ and has a solution if and only if $\det A = 0$. If there is a solution then $sp = rq$ and since $\deg r < \deg p$ clearly one of the factors of p must then be a factor of q. The matrix A appears in the following definition.

Definition 12 *The resultant* $\mathcal{R}(p,q)$ *of* $p(x)$ *and* $q(x)$ *is the* $(m+n)\times(m+n)$ *determinant*

$$\det\begin{pmatrix} a_0 & a_1 & \dots & a_n & 0 & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_n & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & a_0 & a_1 & \dots & a_n \\ b_0 & b_1 & \dots & b_m & 0 & 0 & \dots & 0 \\ 0 & b_0 & b_1 & \dots & b_m & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & b_0 & b_1 & \dots & b_m \end{pmatrix}.$$

Example: A simple example is the condition that a polynomial p has a repeated root α. Then if $p(x)=(x-\alpha)^2 r(x)$, its derivative is $p'(x)=2(x-\alpha)r(x)+(x-\alpha)^2 r'(x)$ so p and p' have a common root hence $\mathcal{R}(p,p')=0$, which is the *discriminant*. For example if $p(x)=ax^2+bx+c$, then

$$\mathcal{R}(p,p')=\det\begin{pmatrix} c & b & a \\ b & 2a & 0 \\ 0 & b & 2a \end{pmatrix}=a(4ac-b^2)$$

which, assuming $a\neq 0$, vanishes if and only if $b^2-4ac=0$.

The key feature to notice about the resultant is that it is a polynomial in the coefficients of p and q which is of degree m in the a_i and degree n in the b_i.

Determinants are often of more theoretical than practical value and there is another way of writing the resultant. Suppose p,q are monic polynomials and let $\lambda_1,\dots,\lambda_n$ be the roots of $p(x)$ and $\mu_1,\dots,\mu_m$ the roots of $q(x)$. Then

$$\prod_{i,j}(\lambda_i-\mu_j)=\prod_i q(\lambda_i)=(-1)^{mn}\prod_j p(\mu_j) \tag{5}$$

clearly vanishes if p and q have a common root. For $p=x^2+bx+c, p'=2x+b$ this gives immediately $c-b^2/4$.

To see the link with the resultant, note that a_k is a homogeneous polynomial in λ_i of degree $n-k$ and b_k a polynomial in μ_j of degree $m-k$. The (i,j) entry in the determinant is therefore of degree $d_{ij}=n+i-j$ if $1\leq i\leq m$ and degree $i-j$ if $m+1\leq i\leq n+m$. The determinant of a $k\times k$ matrix A is a sum

$$\sum_{\sigma\in S_k}\operatorname{sgn}\sigma A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{k\sigma(k)}$$

so in our case each sum is of degree

$$\sum_1^m (n+i-\sigma(i)) + \sum_{m+1}^{m+n}(i-\sigma(i)) = mn + \sum_1^{m+n}(i-\sigma(i)) = mn.$$

So $\mathcal{R}(p,q)$ is a homogeneous polynomial in the $m+n$ variables $(\lambda_1, \ldots \mu_m)$ of degree mn which vanishes precisely when $\lambda_i = \mu_j$ so (5) must be a scalar multiple of $\mathcal{R}(p,q)$, in fact in general

$$\mathcal{R}(p,q) = a_n^m b_m^n \prod_{i,j}(\lambda_i - \mu_j).$$

With this description it is easy to see that

$$\mathcal{R}(p,qr) = \mathcal{R}(p,q)\mathcal{R}(p,r) \tag{6}$$

Now replace $p(x), q(x)$ by $P(x,y,z), Q(x,y,z)$. The resultant is now a polynomial in y and z and we write it as $\mathcal{R}_{P,Q}(y,z)$. Since in

$$P(x,y,z) = a_0(y,z) + a_1(y,z)x + \ldots + a_n(y,z)x^n$$

the coefficient $a_k(y,z)$ is a homogeneous polynomial of degree $n-k$, it follows as we have just seen that $\mathcal{R}_{P,Q}(y,z)$ is a homogeneous polynomial of degree mn.

3.2 Applications

Theorem 8 *Any two algebraic curves in* $\mathbf{P}_2$ *intersect in at least one point.*

Proof: The coefficients a_n, b_m are constants which vanish if and only if $[1,0,0]$ is a point on both curves. Without loss of generality assume then that $a_n \neq 0$, then the resultant $\mathcal{R}_{P,Q}(y,z)$ is a homogeneous polynomial of degree mn so

$$\mathcal{R}_{P,Q}(y,z) = cz^{mn}\prod_i^{nm}(\frac{y}{z} - \nu_i) = c'y^{mn}\prod_i^{nm}(\frac{z}{y} - \nu_i')$$

so there is always a non-zero value $(y,z) = (b,c)$ such that $\mathcal{R}_{P,Q}(b,c) = 0$. Then $P(x,b,c)$ and $Q(x,b,c)$ have a common zero $x = a$ and $[a,b,c] \in \mathbf{P}_2$ lies in both curves. □

Theorem 9 *Two algebraic curves* C, D *in* $\mathbf{P}_2$ *of degrees* n, m *respectively intersect in at most* nm *points, if they have no common component.*

Proof: Suppose a set S of $nm+1$ distinct points lies in $C \cap D$. By a projective transformation assume that $[1,0,0]$ is not one of these, nor does it lie on a line joining any pair of them. Then in particular $P(1,0,0) \neq 0 \neq Q(1,0,0)$ and we can use the resultant: $\mathcal{R}_{P,Q}(y,z)$ is a product of linear factors $bz - cy$ with $(b,c) \neq 0$. Since $\mathcal{R}_{P,Q}(b,c) = 0$ there is an a such that $P(a,b,c) = 0 = Q(a,b,c)$. Conversely if $[a,b,c] \in S$ then $P(a,b,c) = 0 = Q(a,b,c)$ and $(b,c) \neq 0$ since $[1,0,0]$ is not in the intersection, so $bz - cy$ is factor of $\mathcal{R}_{P,Q}(y,z)$.

The points $[a,b,c]$ and $[a',b,c]$ cannot both be in S unless $a' = a$ for then $[1,0,0] = [a'-a, b-b, c-c]$ lies on the line joining them, so (b,c) up to a scalar multiple determines each of the $nm+1$ points of intersection. But if $\mathcal{R}_{P,Q}(y,z)$ is not identically zero, it has at most nm linear factors, which is a contradiction.

If $\mathcal{R}_{P,Q}(y,z) \equiv 0$ then from the properties of the resultant, P and Q have a common factor as polynomials in x over the field of rational functions $p(y,z)/q(y,z)$. But by Gauss' lemma, they have a common polynomial factor, which defines a common component of the curves C and D. □

Theorem 10 *(i) A nonsingular algebraic curve in* $\mathbf{P}_2$ *is irreducible.*

(ii) An irreducible curve in $\mathbf{P}_2$ *has at most a finite number of singular points.*

Proof: (i) Let the reducible curve be defined by $P(x,y,z)Q(x,y,z) = 0$. By Theorem 8 there is a point $[a,b,c]$ where $P(a,b,c) = 0 = Q(a,b,c)$ and differentiating the product PQ at (a,b,c) gives zero.

(ii) Without loss of generality assume $[1,0,0]$ does not lie on the curve C, so that the coefficient of x^n in $P(x,y,z)$ is non-zero. Then $\partial P/\partial x$ is a non-zero homogeneous polynomial of degree $n-1$. C is irreducible and $\partial P/\partial x$ has lower degree so there is no common component, hence by Theorem 9 there are at most $n(n-1)$ points of intersection. The singular points lie amongst these. □

Theorem 11 *(Pascal's mystic hexagram) The pairs of opposite sides of a hexagon inscribed in an irreducible conic meet in three collinear points.*

Proof: Let the successive sides of the hexagon be defined by linear polynomials $L_1, \ldots L_6$, and put $P = L_1L_3L_5$ defining the degree 3 curve C and $Q = L_2L_4L_6$ defining the curve D, then the six vertices of the hexagon lie in $C \cap D$ and also on the conic.

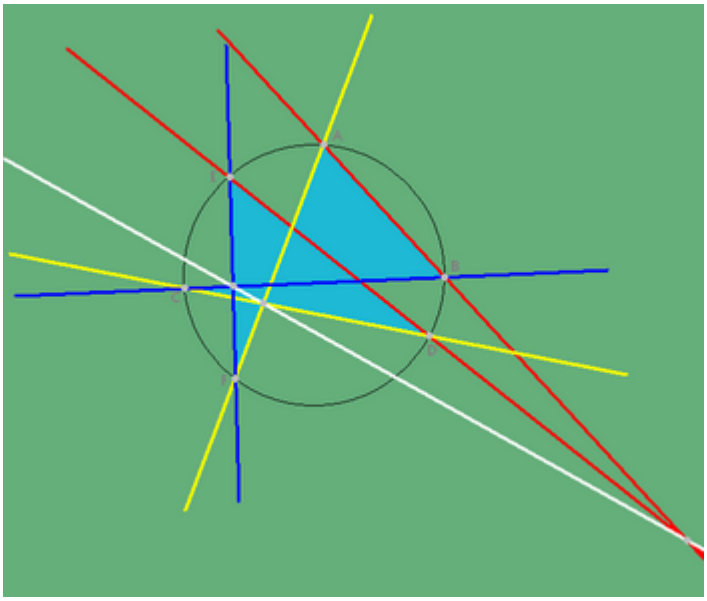
Let $[a,b,c]$ lie on the conic but not on $C \cap D$ then

$$Q(a,b,c)P(x,y,z) - P(a,b,c)Q(x,y,z)$$

defines a degree 3 curve which meets the conic in the six vertices *plus* the point $[a, b, c]$, i.e. in $7 = 2 \times 3 + 1$ points. By Theorem 9, since the conic is irreducible we must have

$$Q(a, b, c)P(x, y, z) - P(a, b, c)Q(x, y, z) = L(x, y, z)R(x, y, z)$$

where $R = 0$ is the equation of the conic. Hence the degree one factor L defines a line which passes through the other $9 - 6$ points of intersection of $C \cap D$. □



3.3 Intersection multiplicity

To count properly we need a definition of the *multiplicity* of the intersection of two curves. For example if we think of a tangent as the limit of a chord where the two points of intersection coalesce into one, then we should count that multiplicity as two. It may also be the case that further points coalesce at the same time, and we need to be able to count these. Or if we intersect two lines L_1, L_2 with a third L, then the two points of intersection with L_1 and L_2 become one if L passes through the singular point $L_1 \cap L_2$.

We use the resultant to make the following (technical) definition:

Definition 13 *The intersection multiplicity $I_p(C, D)$ of curves C and D at p is defined by the following prescription:*

- $I_p(C, D) = \infty$ *if p lies on a common component of C and D.*

- $I_p(C,D) = 0$ *if* p *does not lie on* $C \cap D$
- *if* $p \in C \cap D$*, but lies in no common component, remove any common components to get curves* C', D' *and choose projective coordinates so that* $[1,0,0]$ *does not lie on* $C' \cup D'$*, nor on any line joining distinct points of* $C' \cap D'$*, nor on any tangent line to* C' *or* D' *at a point of* $C' \cap D'$*. Then define* $I_p(C,D)$ *at* $p = [a,b,c]$ *to be the largest integer* k *such that* $(bz - cy)^k$ *divides the resultant* $\mathcal{R}_{P',Q'}(y,z)$*.*

With this definition, we immediately see:

Proposition 12 *Suppose* C *is nonsingular at* p*, and let* T_p *be the tangent line at* p*. Then* $I_p(C,T_p) > 1$*.*

Proof: With the hypotheses of the theorem, $\partial P/\partial x \neq 0$ at $[a,b,c]$, so (writing the partial derivatives as P_x etc.) the resultant, obtained by substituting from the linear equation of the tangent, is a multiple of

$$P(-P_x(p)^{-1}(P_y(p)y + P_z(p)z), y, z).$$

The multiplicity is greater than one if this has a repeated factor, which is where the derivative with respect to y (or z) vanishes. But using the chain rule, differentiating with respect to y at p gives

$$(-P_x^{-1}P_yP_x + P_y)(p) = 0.$$

□

Proposition 13 *If* $p \in C \cap D$ *is a singular point of* C*, then* $I_p(C,D) > 1$*.*

Proof: We can choose coordinates such that $p = [0,0,1]$, so the partial derivatives of P vanish here, hence when we write

$$P(x,y,z) = a_0(y,z) + a_1(y,z)x + \ldots + a_n(y,z)x^n$$

the coefficient $a_0(y,z)$ is divisible by y^2 and $a_1(y,z)$ by y.

Since $Q(0,0,1) = 0$, $b_0(y,z)$ is divisible by y and so

$$b_0(y,z) = b_{01}yz^{m-1} + y^2c_0(y,z) \qquad b_1(y,z) = b_{10}z^{m-1} + yc_1(y,z).$$

If $b_{01} = 0$ then the first column of the resultant is divisible by y^2; if $b_{01} \neq 0$ then only y divides the first column. But taking that factor out and subtracting b_{10}/b_{01} times the first from the second column gives another factor y. So y^2 divides the resultant and the intersection multiplicity is bigger than one. □

The resultant is a fundamentally useful tool but it has the disadvantage that we have to select x from (x, y, z) to define it, and it is not clear that the definition of $I_p(C, D)$ is independent of this choice, so it is useful to know that the following properties, which are clearly independent of choice, actually characterize uniquely $I_p(C, D)$ (see Kirwan's book for details).

Theorem 14 *The intersection multiplicity satisfies the following properties:*

(i) $I_p(C, D) = I_p(D, C)$

(ii) $I_p(C, D) = \infty$ *if* p *lies on a common component of* C *and* D *and is otherwise a nonnegative integer.*

(iii) $I_p(C, D) = 0$ *if and only if* p *does not lie in* $C \cap D$.

(iv) Two distinct lines meet at a point with multiplicity one.

(v) If the curve C *is the union of components* C_1, C_2 *then*

$$I_p(C, D) = I_p(C_1, D) + I_p(C_2, D).$$

(vi) If C *and* D *are defined by* P *and* Q *and* E *is defined by* $PR + Q$*, then*

$$I_p(C, D) = I_p(C, E).$$

Proof:

(i) This is clear since $\mathcal{R}_{P,Q} = \pm\mathcal{R}_{Q,P}$ from (5).

(ii) This was part of the definition.

(iii) If p does lie in the intersection then the resultant has at least a linear factor, so $I_p(C, D) \geq 1$.

(iv) If the lines meet at $p = [0, 0, 1]$ they are given by $ax + by = 0, cx + dy = 0$ so the resultant is $(ad - bc)y$.

(v) This follows from $\mathcal{R}_{P,QR} = \mathcal{R}_{P,Q}\mathcal{R}_{P,R}$.

(vi) If

$$R(x, y, z) = \rho_0(y, z) + \rho_1(y, z)x + \ldots + \rho_{n-m}(y, z)x^{n-m}$$

then the resultant $\mathcal{R}_{P,PR+Q}$ is the determinant of B_{ij} where B_{ij} is equal to the matrix A_{ij} of $\mathcal{R}_{P,Q}$ for $i \leq m$ and to

$$B_{ij} + \sum_{k=i-m}^{i-n} \rho_{i-n-k} A_{kj}$$

for $i > m$. But this is obtaned from A_{ij} by row operations so the determinant is unchanged. □

3.4 Cubic curves

Using the notion of multiplicity, we show here how to reduce a nonsingular cubic curve – defined by a homogeneous polynomial $P(x, y, z)$ of degree 3 – to a standard form. The vector space of such polynomials is 10-dimensional so there are 9 free parameters to specify the cubic curve. A projective transformation is determined by what it does to four points in general position in the projective plane, which gives 8 parameters, so, unlike the conics, we can't expect to reduce each cubic to a single equation. There is one remaining degree of freedom, and we shall prove that

Theorem 15 *After a projective transformation, the equation for any nonsingular cubic curve can be put in the form*

$$y^2z = x(x - z)(x - \lambda z)$$

where $\lambda \neq 0, 1$.

To do this we consider *inflection points*, or flexes.

Definition 14 *A nonsingular point $p \in C$ is called an inflection point if there is a line L through p with $I_p(C, L) \geq 3$. The line is necessarily the tangent to C at p.*

Remark: In calculus inflection points on the graph $y = f(x)$ are defined to be points where the second derivative $f''(x)$ vanishes. To see that this is the same we have to stray from polynomials into holomorphic functions but we shall do this later anyway when we consider algebraic curves as Riemann surfaces. Suppose that $z \neq 0$ then the curve is given by the equation $P(x, y, 1) = 0$ in affine coordinates. If it is nonsingular one of $\partial P/\partial x, \partial P/\partial y$ is non-zero. Suppose it is the latter, then at $(x, y) = (a, b)$ on the curve, the implicit function theorem tells us that there are neighbourhoods V and W of a and b in $\mathbf{C}$ and a holomorphic function $g : V \to W$ such that for $x \in V$ and $y \in W$ $P(x, y, 1) = 0$ if and only if $y = g(x)$.

The tangent line at (a, b) is $(x, y) = (a, b) + t(1, g'(a))$ so (a, b) is an inflection point if and only if t^3 divides $P(a + t, b + tg'(a), 1)$. Now $P(x, g(x), 1) \equiv 0$, and $g(a + t) = b + tg'(a) + t^2g''(a)/2 + t^3h(t)$. Moreover

$$P(x, y, 1) = \sum_{k,\ell} a_{k\ell}x^k y^\ell$$

so

$$\sum_{k,\ell} a_{k\ell}(a+t)^k(b+tg'(a)+t^2g''(a)/2+t^3h(t))^\ell \equiv 0.$$

But then, expanding the power of ℓ,

$$\sum_{k,\ell} a_{k\ell}(a+t)^k(b+tg'(a))^\ell + t^2g''(a)/2\sum_{k,\ell} a_{k\ell}(a+t)^k\ell(b+tg'(a))^{\ell-1} + t^3k(t) \equiv 0.$$

Since, near $(x,y)=(a,b)$,

$$0 \neq \frac{\partial P}{\partial y} = \sum_{k,\ell} a_{k\ell}x^k\ell y^{\ell-1}$$

it follows that t^3 divides the first term, which is $P(a+t, b+tg'(a), 1)$, only where $g''(a)=0$.

In what follows it is convenient to revert to the notation of (x_0,x_1,x_2) for (x,y,z) and to set P_i to be the first derivatives and P_{ij} to be the matrix of second partial derivatives:

$$P_{ij} = \frac{\partial P}{\partial x_i \partial x_j}.$$

Proposition 16 *A nonsingular point $p \in C$ is an inflection point if and only if* $\det P_{ij} = 0$.

Proof: To find the multiplicity of p with respect to a line we have to take the resultant of $P(x,y,z)$ and $ax+by+cz$. But this is just substituting $x=-(by+cz)/a$ into P. It is more convenient to retain the symmetry and consider the line as the set of points

$$[a_0+t\alpha_0, a_1+t\alpha_1, a_2+t\alpha_2]$$

as t varies. Then $I_p(C,L) \geq 3$ if and only if t^3 divides

$$P(a_0+t\alpha_0, a_1+t\alpha_1, a_2+t\alpha_2).$$

Expanding this, we have

$$P(a+t\alpha) = P(a) + t\sum_i P_i(a)\alpha_i + \frac{t^2}{2}\sum_{i,j} P_{ij}(a)\alpha_i\alpha_j + t^3R \tag{7}$$

and t^3 divides this if and only if

$$\sum_i P_i(a)\alpha_i = 0 = \sum_{i,j} P_{ij}(a)\alpha_i\alpha_j.$$

The first equation says that the line is a tangent.

We now use homogeneity – the P_i are homogeneous of degree $n-1$ – so Euler's relation gives

$$(n-1)P_i(a) = \sum_j P_{ij}(a)a_j.$$

This always gives

$$\sum_{i,j} P_{ij}(a)a_ia_j = (n-1)\sum_i P_i(a)a_i = n(n-1)P(a) = 0 \tag{8}$$

and in our case also

$$\sum_{i,j} P_{ij}(a)a_i\alpha_j = (n-1)\sum_i P_i(a)\alpha_i = 0$$

from the first equation. Together with the second equation we see that the quadratic form on the vector space spanned by a and α (the subspace of $\mathbf{C}^3$ defining the line L) vanishes completely. This means that the matrix of the quadratic form with respect to a basis a, α, β is of the form

$$\begin{pmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{pmatrix}$$

and so $\det P_{ij}(a) = 0$.

Conversely take a basis of a,α, β where a and α define the tangent at p. From (8) the matrix is of the form

$$\begin{pmatrix} 0 & 0 & * \\ 0 & * & * \\ * & * & * \end{pmatrix} \tag{9}$$

If $\sum_{ij} P_{ij}a_i\beta_j = 0$ then $\sum_i P_i\beta_i = 0$ by homogeneity, but then $[\beta_0, \beta_1, \beta_2]$ lies on the tangent so a, α, β do not form a basis. Hence the determinant $\det P_{ij}$ vanishes if and only if the central term

$$\sum_{i,j} P_{ij}(a)\alpha_i\alpha_j = 0.$$

□

Remark: The inflection points are the points of $C \cap H$ where H is the curve (called the *Hessian*) with equation $\det P_{ij} = 0$. From Proposition 8 they exist. In fact they are finite in number if the degree of P is bigger than one. To see this, suppose there are infinitely many. Then the Hessian and C have a common component. Since C is nonsingular it must be a component of H and so every point is a flex. But then, using the description of the curve as a graph $y = g(x)$, we see that $g''(x) \equiv 0$ and $g(x) = \alpha + \beta x$. But then

$$P(x, \alpha + \beta x, 1) \equiv 0$$

and $y - \alpha - \beta x$ is a factor of $P(x, y, 1)$ which is a contradiction unless $P(x, y, 1)$ is linear.

Now start the proof of Theorem 15.

Proof: The partial derivatives P_{ij} are of degree $3 - 2 = 1$, so $\det P_{ij} = 0$ is a cubic curve H.

Take $[0, 1, 0]$ to be an inflection point and $z = 0$ to be the tangent. Then from (7) $P(t, 1, 0) = t^3 c$ which means that

$$P(x, y, z) = kx^3 + a_1(x, y)z + a_2(x, y)z^2 + a_3 z^3.$$

Since $z = 0$ is the tangent at $[0, 1, 0]$, $\partial P/\partial z \neq 0$ so the coefficient of $y^2 z$ is nonvanishing. Moreoever clearly the coefficient of y is divisible by z, so the equation can be put in the form

$$y^2 z + yz(\alpha x + \beta z) + b_1(x, z) = 0.$$

Now "complete the square" to write this as $(y - (\alpha x + \beta z)/2)^2 z - b_2(x, z) = 0$ and apply the projective transformation $[x, y, z] \mapsto [x, -\alpha x/2 + y - \beta z/2, z]$ to transform it to $y^2 z = b_3(x, z)$. Since C is nonsingular it is irreducible, so z does not divide $b_3(x, z)$ and therefore the coefficient of x^3 is non-zero so we can write

$$y^2 z = A(x - az)(x - bz)(x - cz)$$

and then $(x, y, z) \mapsto ((x - az)/(b - a), y, z)$ and rescaling y takes it to the form

$$y^2 z = x(x - z)(x - \lambda z).$$

If $\lambda = 0$ then $[0, 0, 1]$ is a singularity and if $\lambda = 1$ then $[1, 0, 1]$ is. □

3.5 Bézout's theorem

Now that we have a definition of intersection multiplicity, we can formulate the basic theorem about the intersection of curves:

Theorem 17 *(Bézout's theorem) If C and D are two algebraic curves in $\mathbf{P}_2$ of degrees n, m with no common component, then*

$$\sum_{p \in C \cap D} I_p(C, D) = mn.$$

Proof: Using the coordinates in the definition of multiplicity, we express the resultant as a product of linear factors:

$$\mathcal{R}_{P,Q}(y, z) = \prod_i (c_i z - b_i y)^{e_i}$$

where $e_1 + \ldots + e_k = mn$.

By the arguments in Theorems 8 and 9 each such factor gives a point $p_i \in C \cap D$ with $I_{p_i}(C, D) = e_i$. □

When is mn the actual number of intersections? Clearly only when $I_{p_i}(C, D) = 1$. We need to the following:

Proposition 18 *The intersection multiplicity $I_p(C, D)$ is equal to one if and only if p is a nonsingular point of C and D and the tangent lines to C and D at p are distinct.*

Proof: From Proposition 13, if $I_p(C, D) = 1$, p must be a nonsingular point of C and D. As usual, choose coordinates such that $p = [0, 0, 1]$. We need to show that the tangent lines coincide if and only if y^2 divides the resultant $\mathcal{R}_{P,Q}(y, z)$, or equivalently that the derivative of $\mathcal{R}_{P,Q}(y, 1)$ vanishes at $y = 0$.

Now by assumption $[1, 0, 0]$ does not lie on the tangent line to p for either curve so $\partial P/\partial x(0, 0, 1) \neq 0$ and similarly for Q. The implicit function theorem then tells us that in a suitable small neighbourhood, the solution x of $P(x, y, 1) = 0$ is a holomorphic function of y. In other words, near $[0, 0, 1]$, the roots $\lambda_1(y), \mu_1(y)$ of P and Q which coincide when $y = 0$ are holomorphic functions of y.

Thus

$$P(x, y, 1) = (x - \lambda_1(y))\ell(x, y) \qquad Q(x, y, 1) = (x - \mu_1(y))m(x, y)$$

for polynomials ℓ, m in x with coefficients which are holomorphic functions of y. Then the resultant $\mathcal{R}_{P,Q}(y,1) = (\lambda_1(y) - \mu_1(y))S(y)$ where $S(y)$ is holomorphic. Differentiating at $y = 0$,

$$\frac{\partial \mathcal{R}_{P,Q}(y,1)}{\partial y}|_{y=0} = (\lambda_1'(0) - \mu_1'(0))S(0). \tag{10}$$

We shall show next that $S(0) \neq 0$.

Since $\partial P/\partial x(0,0,1) \neq 0$, $x = 0$ is not a repeated root of $P(x,0,1)$ so for $i \neq 1$, $\lambda_i(0) \neq 0$ and similarly for Q. If $\lambda_i(0) = \mu_j(0)$ for $i, j > 1$ then $[0,0,1]$ and $[\lambda_i(0),0,1]$ are distinct points in $C \cap D$ and $[1,0,0]$ lies on the line joining them which contradicts our assumptions. Now $S(y)$ is a product of resultants and we see here that there is no other coincidence of roots than $\lambda_1(0) = \mu_1(0)$ at $y = 0$. Thus $S(0) \neq 0$.

It follows that the derivative in equation (10) vanishes if and only if $\lambda_1'(0) - \mu_1'(0) = 0$. Now since $P(\lambda_1(y), y, 1) \equiv 0$, differentiating with respect to y gives

$$\frac{\partial P}{\partial x}\lambda_1'(y) + \frac{\partial P}{\partial y} = 0$$

and at $[0,0,1]$ by Euler's identity $\partial P/\partial z = nP = 0$, so the tangent line to C is $x - \lambda_1'(0)y = 0$ and for D $x - \mu_1'(0)y = 0$. Hence the tangents coincide if and only if $\lambda_1'(0) - \mu_1'(0) = 0$, which proves the theorem.

□

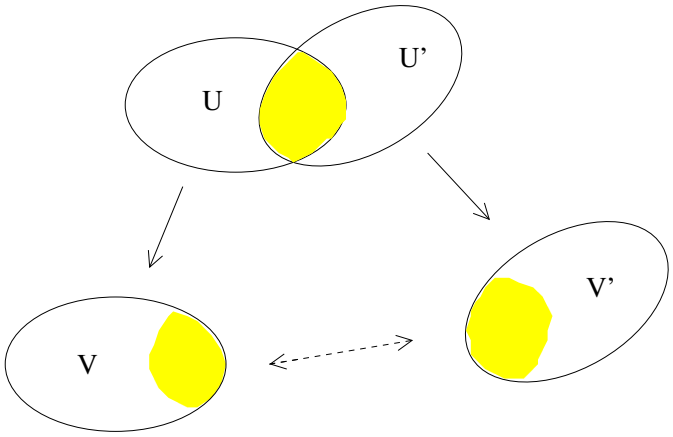
4 The genus of a curve

4.1 Riemann surfaces

Nonsingular projective algebraic curves provide us with a rich source of examples of Riemann surfaces, objects dealt with in the *Geometry of Surfaces* course. This part of the notes repeats some of the facts in the notes for the earlier course, and applies them to algebraic curves. Let us recall the notion of an abstract surface: each point has a neighbourhood U and a homeomorphism φ_U from U to an open set V in $\mathbf{R}^2$. If two such neighbourhoods U, U' intersect, then

$$\varphi_{U'}\varphi_U^{-1} : \varphi_U(U \cap U') \to \varphi_{U'}(U \cap U')$$

is a homeomorphism from one open set of $\mathbf{R}^2$ to another.



If we identify $\mathbf{R}^2$ with the complex numbers $\mathbf{C}$ then we can define:

Definition 15 *A Riemann surface is a surface with a class of homeomorphisms φ_U such that each map $\varphi_{U'}\varphi_U^{-1}$ is a* holomorphic *homeomorphism.*

We call each function φ_U a holomorphic coordinate. If we compose with an invertible holomorphic map $f : \varphi_U(U) \to \mathbf{C}$ then $f \circ \varphi_U$ is another holomorphic coordinate. We shall be studying properties of Riemann surfaces which do not depend on a particular choice of coordinate.

Examples:

1. Let X be the complex projective line $X = \mathbf{P}_1$. Let $U = \{[z_0, z_1] \in \mathbf{P}_1 : z_0 \neq 0\}$ with $\varphi_U(z) = z_1/z_0 \in \mathbf{C}$. Now take $U' = \{[z_0, z_1] \in \mathbf{P}_1 : z_1 \neq 0\}$ and define

$\varphi_U(z) = z_0/z_1 \in \mathbf{C}$. Then

$$\varphi_U(U \cap U') = \mathbf{C}\backslash\{0\}$$

and

$$\varphi_U \varphi_{U'}^{-1}(z) = z^{-1}$$

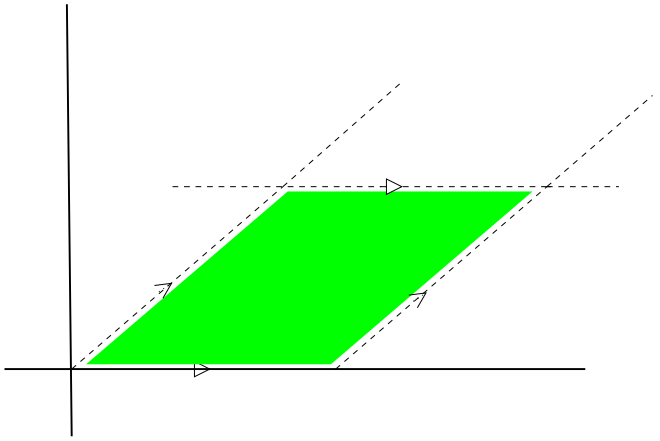
which is holomorphic.

2. Let $\omega_1, \omega_2 \in \mathbf{C}$ be two complex numbers which are linearly independent over the reals, and define an equivalence relation on $\mathbf{C}$ by $z_1 \sim z_2$ if there are integers m, n such that $z_1 - z_2 = m\omega_1 + n\omega_2$. Let X be the set of equivalence classes (with the quotient topology). A small enough disc V around $z \in \mathbf{C}$ has at most one representative in each equivalence class, so this gives a local homeomorphism to its projection U in X. If U and U' intersect, then the two coordinates are related by a map

$$z \mapsto z + m\omega_1 + n\omega_2$$

which is holomorphic.

This surface is topologically described by noting that every z is equivalent to one inside the closed parallelogram whose vertices are $0, \omega_1, \omega_2, \omega_1 + \omega_2$, but that points on the boundary are identified:



We thus get a torus this way. Another way of describing the points of the torus is as the quotient group $\mathbf{C}/\Gamma$ where Γ is the subgroup consisting of complex numbers of the form $m\omega_1 + n\omega_2$.

3. Let C be a nonsingular projective algebraic curve defined by $P(x, y, z) = 0$. Every point lies in an affine open set of $\mathbf{P}_2$ which is homeomorphic to $\mathbf{C}^2$. On $z \neq 0$ its equation is $P(x, y, 1) = 0$ and if C is nonsingular one of $\partial P/\partial x, \partial P/\partial y$ is non-zero. Suppose it is the latter, then at $(x, y) = (a, b)$ on the curve, the implicit function

theorem tells us that there are neighbourhoods V and W of a and b in $\mathbf{C}$ and a holomorphic function $g : V \to W$ such that for $x \in V$ and $y \in W$ $P(x, y, 1) = 0$ if and only if $y = g(x)$. Hence for $(x, y) \in C \cap (V \times W)$ the function x has an inverse $x \mapsto (x, g(x))$, and this is a local coordinate for C.

If $\partial P/\partial x$ is non-vanishing we can do the same interchanging the roles of x and y, and get $x = h(y)$. Where both are non-vanishing

$$y = g(h(y))$$

and we have an invertible holomorphic function relating the two local coordinates.

On the affine set $y \neq 0$, the equation of the curve is $P(\tilde{x}, 1, \tilde{z})$ where, when $z \neq 0$, $\tilde{x} = x/y, \tilde{z} = 1/y$ and it is easy to see that the holomorphic coordinates on the intersection of these two open sets is holomorphic and invertible.

Definition 16 *A holomorphic map between Riemann surfaces X and Y is a continuous map $f : X \to Y$ such that for each holomorphic coordinate φ_U on U containing x on X and ψ_W defined in a neighbourhood of $f(x)$ on Y, the composition*

$$\psi_W \circ f \circ \varphi_U^{-1}$$

is holomorphic.

Example: Consider a projective algebraic curve C defined by $P(x, y, z)$. If $[0, 0, 1] \neq C$ we can define a map

$$f : C \to \mathbf{P}_1$$

by

$$f([x, y, z]) = [x, y]. \tag{11}$$

This is well defined because x and y are not simultaneously zero, and will play an important role in what follows.

4.2 Maps to $\mathbf{P}_1$

Maps from an algebraic curve to the projective line have another interpretation:

Definition 17 *A meromorphic function f on a Riemann surface X is a map to $\mathbf{C} \cup \{\infty\}$ such that for each coordinate neighbourhood $f \circ {\varphi_U}^{-1}$ is a meromorphic function on $\varphi_U(U) \subseteq \mathbf{C}$.*

Examples: 1. A rational function

$$f(z) = \frac{p(z)}{q(z)}$$

where p and q are polynomials is a meromorphic function on $\mathbf{P}_1$.

2. Look at the map (11) where $f = y/x$. In any of our local coordinates this is the ratio of two holomorphic functions and so is meromorphic. We could also take any polynomial in f, or ratios of these.

This is an analytic view of meromorphic functions: we can add, multiply them together and also divide by non-zero functions. These are all meromorphic and form a field. On the other hand, any individual meromorphic function on X can be geometrically represented as a map to $\mathbf{P}_1$: if we remove $f^{-1}(\infty)$, then f is just a holomorphic function F with values in $\mathbf{C}$, so on this part of X, we define the map

$$x \mapsto [F(x), 1] \in \mathbf{P}_1.$$

If $f(x) = \infty$, and U is a small enough coordinate neighbourhood of x, then $(f \circ \varphi_U^{-1})^{-1}$ is a holomorphic function $\tilde{F}$, and the map is defined by $x \mapsto [1, \tilde{F}(x)] \in \mathbf{P}_1$.

Before proceeding, recall some basic facts about ordinary holomorphic functions (see, for example, *Introduction to Complex Analysis*, Second Edition, H. A. Priestley, OUP, Price: £23.00 (Paperback)):

- A holomorphic function has a convergent power series expansion in a neighbourhood of each point at which it is defined:

  $$f(z) = a_0 + a_1(z - c) + a_2(z - c)^2 + \ldots$$

- If f vanishes at c then

  $$f(z) = (z - c)^m(a_0 + a_1(z - c) + \ldots)$$

  where $a_0 \neq 0$. In particular zeros are isolated.

- If f is non-constant it maps open sets to open sets.

- If $f'(c) \neq 0$ then f has a local holomorphic inverse g (this is very useful for changing a local coordinate to a more convenient one.)

- $|f|$ cannot attain a maximum at an interior point of a disc ("maximum modulus principle") unless f is constant.
- $f : \mathbf{C} \to \mathbf{C}$ preserves angles between differentiable curves, both in magnitude and sense.

Remark: One consequence of the maximum modulus principle is that a holomorphic map $f : X \to \mathbf{C}$ from a compact connected Riemann surface X must be a constant: by compactness $|f|$ has a maximum at a, so in some coordinate neighbourhood of a is represented by a holomorphic function F with an interior maximum modulus. It follows that $f(x) = c$ in a neighbourhood. Now consider the set of all x such that $f(x) = c$ in a neighbourhood of x. This is open and non-empty, but is also closed because the zeros of $f(x) - c$ are isolated.

Now return to the situation of a holomorphic map $f : X \to \mathbf{P}_1$. For each point, this is described by a locally defined holomorphic function $F = f \circ \varphi_U^{-1}$.

If the inverse image of $a \in \mathbf{P}_1$ is infinite, then it has a limit point x by compactness of X. In a holomorphic coordinate z around x with $z(x) = 0$, f is defined by a holomorphic function F with a sequence of points $z_n \to 0$ for which $F(z_n) - a = 0$. But the zeros of a holomorphic function are isolated, so we deduce that $f^{-1}(a)$ is a finite set.

Example: Take the map (11). The inverse image of $a = [a_0, a_1]$ is the set of points $[a_0 t, a_1 t, z]$ such that $P(a_0 t, a_1 t, z) = 0$ which is a homogeneous polynomial in t and z of degree n and splits into $\leq n$ factors.

By a similar argument the points at which the derivative F' vanishes are finite in number (check using the chain rule that this condition is independent of the holomorphic coordinate).

Definition 18 *Let $f : X \to Y$ be a holomorphic map of Riemann surfaces. The point $x \in X$ is a ramification point if in local coordinates f is represented by a holomorphic function F such that $F' = 0$ at x.*

If f is any holomorphic function on $\mathbf{C}$ such that $f'(0) = 0$, we have

$$f(z) = z^n(a_0 + a_1 z + \ldots)$$

with $a_0 \neq 0$. We can expand

$$(a_0 + a_1 z + \ldots)^{1/n} = a_0^{1/n}(1 + b_1 z + \ldots)$$

in a power series and define

$$w = a_0^{1/n} z(1 + b_1 z + \ldots).$$

Since $w'(0) \neq 0$ we can think of w as a new coordinate and then the map becomes simply

$$w \mapsto w^n.$$

So, thinking geometrically of $\mathbf{P}_1$ as a Riemann surface where we are allowed to change coordinates, a ramification point can be locally put in the form $z \mapsto z^n$. The integer n is its *ramification index*. If F' is not zero at $z = 0$ then clearly the index is 1.

Proposition 19 *In the map (11), the ramification points are those points $p \in C$ at which the tangent line T_p at p passes through $[0, 0, 1]$. The ramification index is the intersection multiplicity $I_p(C, T_p)$.*

Proof: Assume that $\partial P/\partial y \neq 0$ and use x as a local coordinate to represent the curve as $y = g(x)$. The map is then $g(x)/x$ if $x \neq 0$ and $x/g(x)$ if $x = 0$. Assume the first case, then $F' = 0$ if and only if $xg'(x) - g(x) = 0$. But $P(x, g(x), 1) = 0$ so

$$\frac{\partial P}{\partial x} + g'(x)\frac{\partial P}{\partial y} = 0.$$

At the ramification point $y = g(x) = xg'(x)$ and $g'(x) \neq 0$ so

$$x\frac{\partial P}{\partial x} + y\frac{\partial P}{\partial y} = 0$$

but from Euler's identity this means that $\partial P/\partial z = 0$. Hence the tangent line passes through $[0, 0, 1]$.

In coordinates suppose the ramification point is $x = c$ and $g(c)/c = a$ then the tangent line is $y = ax$. If the ramification index is n, then locally

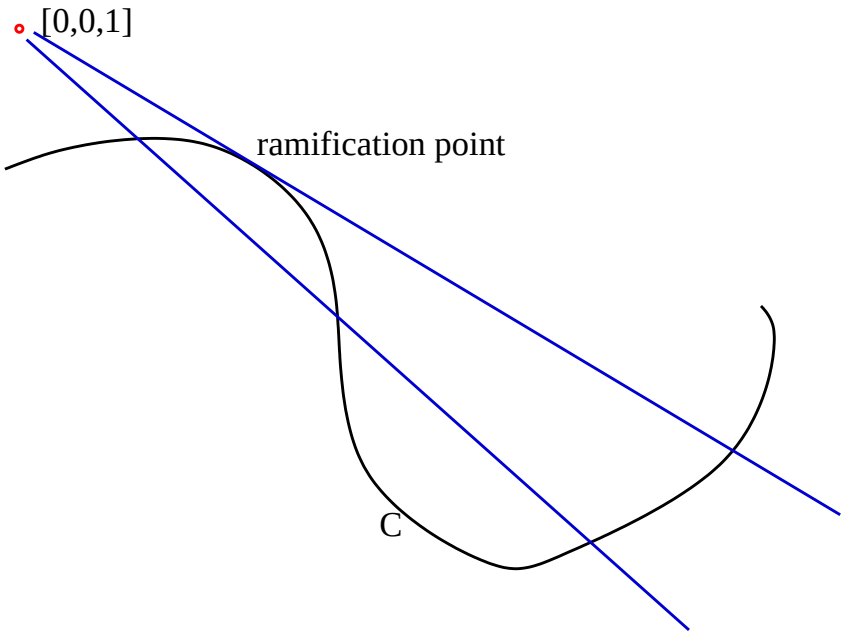
$$g(x) = ax + (x - c)^n h(x)$$

where $h(c) \neq 0$. Then since $P(x, g(x), 1) \equiv 0$ putting $x - c = s$,

$$P(c + s, ac + as + s^n k(s), 1) \equiv 0.$$

As in Proposition 16 this means that $P(c + s, ac + as, 1)$ is divisible by s^n and no higher power, so the intersection multiplicity with the tangent $y = ax$ is n. □

Remark: The map in (11) has a more geometric character: for each point $p \in C$ consider the line joining it to $[0,0,1]$. This defines a map to the space of lines through $[0,0,1]$, but this is a projective line, though not a line in the $\mathbf{P}_2$ in which C lies. Instead, if we consider the dual space V^* of linear maps $f : V \to \mathbf{C}$, those which annihilate $(0,0,1)$ form a two-dimensional subspace, hence a line in $P(V^*)$, the space of all lines in $P(V)$ through $[0,0,1]$.



4.3 The degree-genus formula

From the *Geometry of Surfaces* course we have:

Proposition 20 *A Riemann surface is orientable.*

and hence a nonsingular projective algebraic curve is a compact orientable surface.

We also have:

Proposition 21 *A nonsingular projective algebraic curve is connected.*

Proof: First consider the special curve $x^n + y^n - z^n = 0$ (sometimes called a *Fermat curve* for obvious reasons). The intersection multiplicity of $y - z = 0$ with the curve

is clearly n, so $[0,1,1]$ is a ramification point of the map $f = y/z$ with multiplicity n. This means that $f^{-1}(U)$ is connected for a small neighbourhood U of $1 \in \mathbf{C}$. If there is another connected component C_0, then $1 \neq f(C_0)$ but then f maps C_0 to $\mathbf{C} \cong \mathbf{P}_1 \setminus \{1\}$ and so f is a constant c. But then the line $y - cz = 0$ divides P but we have assumed that C is nonsingular. So this curve is definitely connected.

Intuitively, it is clear that if we change the coefficients slightly, the number of connected components doesn't change. Below we will give a better justification of this fact, but let's assume that for the moment. The condition for nonsingularity is the vanishing of a polynomial in the coefficients so if we take curves define by P and Q, then $tP(x,y,z) + (1-t)Q(x,y,z)$ for $t \in \mathbf{C}$ will be nonsingular unless a polynomial in t vanishes at a finite number of points or is identically zero. If the latter we can replace this path between P and Q by a series of such complex "intervals" for which the singular curves are given by the vanishing of a polynomial in t. Either way, we can avoid a finite number of points in $\mathbf{C}$ by a *real* path joining P to Q, and so have a path of curves all of which are nonsingular. If we start with $P(x,y,z) = x^n + y^n - z^n$, then since $P = 0$ is connected, so is the curve defined by Q. □

Remark: We give here a bit more detail about one way of seeing that if we vary the coefficients continuously through nonsingular curves, then in fact any two such curves are *homeomorphic*, much stronger than just having the same number of connected components. It involves the first fundamental form for a three-dimensional object – a differentiable 3-manifold – but it is just a generalization of what you have seen in the *Geometry of Surfaces* course.

So suppose $P(x,y,z,t)$ is a homogeneous polynomial whose coefficients depend differentiably on the real parameter t, and so that for each t the curve $P(x,y,z,t) = 0$ is nonsingular. Inside $\mathbf{P}_2 \times \mathbf{R}$ we look at the set $P(x,y,z,t) = 0$. In affine coordinates, if $P_x(x,y,1,t) \neq 0$ then by the implicit function theorem with $y = u + iv$, we have a locally defined function $x(u,v,t)$ such that

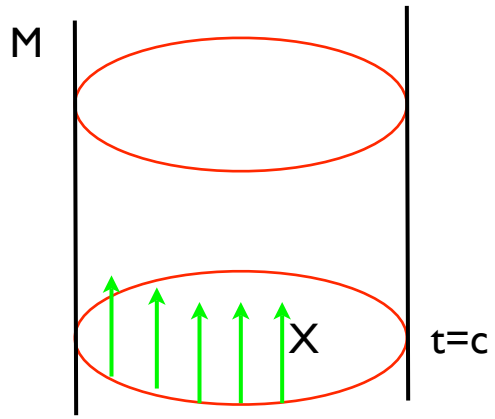
$$P(x(u,v,t), u+iv, 1, t) \equiv 0.$$

Repeating for the other affine open sets means that the set $P(x,y,z,t) = 0$ has the structure of a 3-dimensional differentiable manifold M, with local coordinates (u,v,t). Moreover, t is a well-defined smooth function on M and since t is always part of a coordinate system, its derivative never vanishes. The level set $t = c$ is the algebraic curve (or real surface in this context), $P(x,y,z,c) = 0$.

We now want to introduce a Riemannian metric, or first fundamental form on M. This means measuring the length of smooth curves on $M \subset \mathbf{P}_2 \times \mathbf{R}$. there are many

ways of doing this, the easiest is to embed $\mathbf{P}_2 \times \mathbf{R}$ in some Euclidean space $\mathbf{R}^N$ and just use the usual length of a curve there. Since a point in $\mathbf{P}_2$ is a one-dimensional subspace of $\mathbf{C}^3$ we can describe it by a 3×3 matrix which acts as 1 in this direction and -1 on the orthogonal complement with respect to the Hermitian inner product on $\mathbf{C}^3$. The space of all 3×3 complex matrices has real dimension 18, so this gives an embedding $\mathbf{P}_2 \times \mathbf{R} \subset \mathbf{R}^{19}$.

Using this inner product on tangent vectors to M, we take the one-dimensional orthogonal complement to the tangent space of the surface $t = c$. Since t is a constant on $t = c$ and its derivative is non-zero, then the derivative of t in this normal direction is non-zero, so we can find a family X of tangent vectors to M, normal to $t = c$, such that the derivative of t in this direction is 1. Vary c and we get a vector field on M always pointing normal to the level sets of t and with the property that the derivative of t in this direction is 1.



We now integrate this vector field for small values of t. Analytically this means solving a differential equation. If we think of X as a wind velocity, we want to see where the surface $t = c$ gets blown to after time s. In local coordinates this is an equation of the form

$$\frac{du}{ds} = a(u, v, t) \quad \frac{dv}{ds} = b(u, v, t) \quad \frac{dt}{ds} = 1$$

where the last relation holds since the derivative of t in the direction X is one. Since $t = c$ is compact, the existence theorems for differential equations say that for s small we have a solution and this means that $(u(0), v(0), t(0)) \mapsto (u(s), v(s), t(s))$ gives a diffeomorphism from $t = c$ to $t(s) = c + s$. By connectedness this extends for t lying in a whole interval.

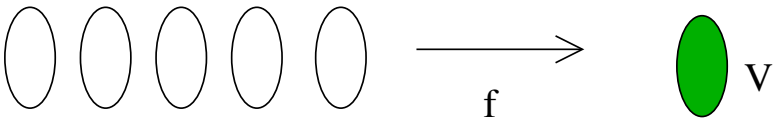
A nonsingular algebraic curve is thus a compact connected oriented surface which means that C is determined up to homeomorphism by its Euler characteristic, which is of the form $\chi(C) = 2 - 2g$.

Definition 19 *The genus of a nonsingular algebraic curve C is the integer $g \geq 0$ such that $\chi(C) = 2 - 2g$.*

Example: The projective line $\mathbf{P}_1 = \mathbf{C} \cup \{\infty\}$ is homeomorphic to a sphere by stereographic projection. It has genus zero.

Using the map to $\mathbf{P}_1$ considered above, we shall calculate g in terms of the degree n of the polynomial $P(x, y, z)$ which defines C, by using the Riemann-Hurwitz formula.

Recall that for a map $f : C \to \mathbf{P}_1$ there are two types of points: if $F'(x) \neq 0$, then the inverse function theorem tells us that f maps a neighbourhood U_x of $x \in C$ homeomorphically to a neighbourhood V_x of $f(x) \in \mathbf{P}_1$. If $F'(x) = 0$ then the map looks like $z \mapsto z^n$ which is not a homeomorphism – the inverse image of 0 is a single point but of any other nearby point it is n points. Define V to be the intersection of the V_x as x runs over the finite set of points such that $f(x) = a$, then $f^{-1}V$ consists of a finite number of disjoint open sets, each one of which is mapped to V by a map of one of the above forms.



Removing the finite number of images under f of ramification points (this is called the *branch locus* of the map) we get a sphere minus a finite number of points. This is connected. The number of points in the inverse image of a point in this punctured sphere is integer-valued and continuous, hence constant. It is called the *degree* d of the map f.

Example: For the map (11) the inverse image of $a = [a_0, a_1]$ is the set of points $[a_0 t, a_1 t, z]$ such that $P(a_0 t, a_1 t, z) = 0$. This inverse image contains no ramification points if this has no repeated factors, so the degree of the map is the degree of the polynomial $P(x, y, z)$.

Now recall from the *Geometry of Surfaces* course:

Theorem 22 *(Riemann-Hurwitz) Let $f : X \to \mathbf{P}_1$ be a holomorphic map of degree d on a closed connected Riemann surface X, and suppose it has ramification points $x_1, \ldots, x_n$ of multiplicity m_k. Then*

$$\chi(X) = 2d - \sum_{k=1}^{n} (m_k - 1)$$

Proof: The idea is to take a triangulation of the sphere $\mathbf{P}_1$ such that the image of the ramification points are vertices (see Kirwan's book for more details). Now take a finite subcovering of $\mathbf{P}_1$ by open sets of the form V above where the map f is either a homeomorphism or of the form $z \mapsto z^m$. Subdivide the triangulation into smaller triangles such that each one is contained in one of the sets V. Then the inverse images of the vertices and edges of $\mathbf{P}_1$ form the vertices and edges of a triangulation of X.

If the triangulation of $\mathbf{P}_1$ has V vertices, E edges and F faces, then clearly the triangulation of X has dE edges and dF faces. It has fewer vertices, though — in a neighbourhood where f is of the form $w \mapsto w^m$ the origin gives a single vertex instead of m of them. For each ramification point of order m_k we therefore have one vertex instead of m_k. The count of vertices is therefore

$$dV - \sum_{k=1}^{n} (m_k - 1).$$

Thus

$$\chi(X) = d(V - E + F) - \sum_{k=1}^{n} (m_k - 1) = 2d - \sum_{k=1}^{n} (m_k - 1)$$

using $\chi(\mathbf{P}_1) = 2$. □

Clearly the argument works just the same for a holomorphic map $f : X \to Y$ and then

$$\chi(X) = d\chi(Y) - \sum_{k=1}^{n} (m_k - 1).$$

We can now calculate the genus:

Theorem 23 *Let C be a non-singular projective algebraic curve of degree n. Then the genus of C is*

$$g = \frac{(n-1)(n-2)}{2}.$$

Proof: Take our familiar map $f = y/x$. As we saw above, the ramification points occur where the tangents pass through $[0,0,1]$ and are therefore given by the equation $\partial P/\partial z = 0$. The multiplicity is bigger than 2 only if $I_p(C, T_p) > 2$, i.e. if p is an inflection point, but there are only finitely many of these, so by a projective transformation we can assume that $[0,0,1]$ does not lie on the tangent to any one of them. This means that each m_k in the Riemann-Hurwitz formula is 2, and it remains to calculate the number of ramification points.

This is the number of points of intersection of $P = 0$, the curve C of degree n, and $\partial P/\partial z = 0$, a curve D which is of degree $n-1$. Since C is nonsingular it is irreducible, and so C and D can have no common component. We will use Bézout's theorem, so we need to check that $[a] = [a_0, a_1, a_2] \in C \cap D$ is a nonsingular point of D and that the tangent lines are distinct. Now (P_{zx}, P_{zy}, P_{zz}) is not identically zero at $[a_0, a_1, a_2]$ because this would make the Hessian of C vanish and we know that $[a_0, a_1, a_2]$ is not an inflection point. This shows that D is nonsingular here.

Suppose that the tangents of C and D coincide then (P_{zx}, P_{zy}, P_{zz}) is a multiple of (P_x, P_y, P_z). As in our discussion of inflection points we use the symmetric bilinear form B defined by the matrix of partial derivatives P_{ij}. Then $B(a,a) = 0 = B(a,\alpha)$ where the tangent line joins $[a]$ and $[\alpha]$. Put $v = (0,0,1)$.

By the Euler identity

$$a_0 P_{zx} + a_1 P_{zy} + a_2 P_{zz} = (n-1)P_z = 0$$

since $P_z(a) = 0$. This gives $B(a,v) = 0$. Moreover since $P_{zz}(a) = \lambda P_z(a) = 0$, we have $B(v,v) = 0$.

Since $[a]$ is not an inflection point, $\det B \neq 0$ so from

$$0 = B(a,a) = B(a,\alpha) = B(a,v)$$

we deduce $v = \mu a + \nu\alpha$. But then

$$0 = B(v,v) = \nu^2 B(\alpha,\alpha)$$

and, as in Proposition 16, this gives $\det B = 0$ unless $\nu = 0$. But then $[a] = [0,0,1]$ which we have specifically excluded. We conclude that the tangents are distinct and

it follows that the conditions for Bézout's theorem hold, so the number of ramification points is exactly $n(n-1)$.

From the Riemann-Hurwitz formula we obtain

$$2 - 2g = 2n - n(n-1)$$

and so

$$g = \frac{1}{2}(n-1)(n-2).$$

□

Remark: When $n = 1$, $g = 0$ and we have already seen that the projective line is homeomorphic to the sphere. The genus also vanishes when $n = 2$, a conic, but from Theorem 7, we saw that this is homeomorphic to the sphere too.

4.4 The torus and the cubic

When $n = 3$ we get $g = 1$ which means that a nonsingular cubic surface is homeomorphic to a torus. Here is a more concrete realization of that fact. We take the Riemann surface which is the second example in Examples 4.1. This is the set of equivalence classes of $z \in \mathbf{C}$ where $z_1 \sim z_2$ if there are integers m, n such that $z_1 - z_2 = m\omega_1 + n\omega_2$ and is clearly a torus.

As in the *Geometry of Surfaces* course, define

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \neq 0} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right)$$

where the sum is over all non-zero $\omega = m\omega_1 + n\omega_2$, with m, n integers. Since for $2|z| < |\omega|$

$$\left| \frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right| \leq 10 \frac{|z|}{|\omega|^3}$$

this converges uniformly on compact sets so long as

$$\sum_{\omega \neq 0} \frac{1}{|\omega|^3} < \infty.$$

But $m\omega_1 + n\omega_2$ is never zero if m, n are real so we have an estimate

$$|m\omega_1 + n\omega_2| \geq k\sqrt{m^2 + n^2}$$

so by the integral test we have convergence. Because the sum is essentially over all equivalence classes, we have

$$\wp(z + m\omega_1 + n\omega_2) = \wp(z)$$

so that this is a meromorphic function on the surface X. It is called the Weierstrass $\wp$-function.

We now want to know geometrically what this map from a torus to the projective line looks like.

Firstly, $\wp$ has degree 2 since $\wp(z) = \infty$ only at $z = 0$ and there it has multiplicity 2. The multiplicity of any ramification point cannot be bigger than this because then it will look like $z \mapsto z^n$ and a non-zero point will have at least n inverse images. Thus the only possible value at the ramification points in this example is $m_k = 2$. The Riemann-Hurwitz formula gives:

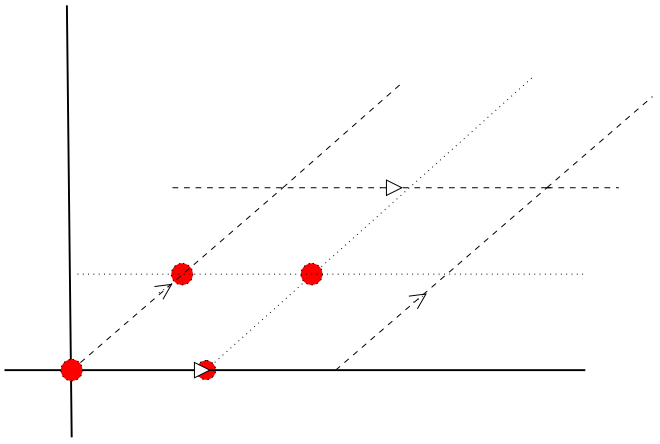
$$0 = 4 - n$$

so there must be exactly 4 ramification points. In fact we can see them directly, because $\wp(z)$ is an even function, so the derivative vanishes if $-z = z$. Of course at $z = 0$, $\wp(z) = \infty$ so we should use the other coordinate on S: $1/\wp$ has a zero of multiplicity 2 at $z = 0$. To find the other points recall that $\wp$ is doubly periodic so $\wp'$ vanishes where

$$z = -z + m\omega_1 + n\omega_2$$

for some integers m, n, and these are the four points

$$0, \omega_1/2, \omega_2/2, (\omega_1 + \omega_2)/2 :$$



The geometric Riemann-Hurwitz formula has helped us here in the analysis by showing us that the only zeros of $\wp'$ are the obvious ones. Suppose, by a projective

transformation of $\mathbf{P}_1$, that their images are ∞ and the three finite points e_1, e_2, e_3 where

$$e_1 = \wp(\omega_1/2), \quad e_2 = \wp(\omega_2/2), \quad e_3 = \wp((\omega_1+\omega_2)/2).$$

The derivative $\wp'(z)$ vanishes at three points, each with multiplicity 1. At each of these points $\wp$ has the local form

$$\wp(z) = e_1 + (z - \omega_1/2)^2(a_0 + \ldots)$$

and so

$$\frac{1}{\wp'(z)^2}(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3)$$

is a well-defined holomorphic function on X away from $z = 0$. But $\wp(z) \sim z^{-2}$ near $z = 0$, and so $\wp'(z) \sim -2z^{-3}$ and hence this function is finite at $z = 0$ with value $1/4$. By the maximum argument, since X is compact, the function is a constant, namely $1/4$, and

$$\wp'(z)^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3). \tag{12}$$

This is the equation of a cubic curve C:

$$y^2 = 4(x - e_1)(x - e_2)(x - e_3).$$

Now $\wp : X \to \mathbf{P}_1$ is surjective (otherwise the degree would be zero!), so $\wp(z)$ takes every value in $\mathbf{P}_1$. Moreover, since $\wp(-z) = \wp(z)$, $\wp'(-z) = -\wp'(z)$ so for each value of x there is a z for both values of y.

Therefore $z \mapsto [\wp(z), \wp'(z), 1]$ defines a homeomorphism from X to C.

5 The Riemann-Roch theorem

5.1 Divisors

Let f be a meromorphic function on an open set $U \subseteq \mathbf{C}$. If $f(a) = 0$ then near a

$$f(z) = (z-a)^m g(z)$$

where $m > 0$ is the multiplicity of the zero and $g(z)$ is holomorphic with $g(a) \neq 0$. So if $\tilde{f}$ has a zero at a with the same multiplicity then

$$\frac{\tilde{f}}{f} = \frac{\tilde{g}(z)}{g(z)}$$

which is holomorphic. The multiplicity is independent of the coordinate z.

Similarly near a pole b

$$f(z) = (z-b)^{-n} h(z)$$

where $h(b) \neq 0$.

Now consider a meromorphic function f on an algebraic curve C. Suppose its zeros are $p_1, \ldots, p_k$ with multiplicities $m_1, \ldots, m_k$ and its poles are $q_1, \ldots, q_\ell$ with multiplicities $n_1, \ldots, n_\ell$. If $\tilde{f}$ is another meromorphic function with exactly the same zeros and poles and multiplicities, then $\tilde{f}/f$ is a (non-vanishing) holomorphic function and by the maximum principle this is a constant. So, up to a constant multiple, f is determined by the zeros, poles and multiplicities. We write this is

$$\sum_{i=1}^{k} m_k p_i - \sum_{i=1}^{\ell} n_i q_i.$$

Definition 20 *A divisor D on a curve C is a formal sum*

$$D = \sum_{p \in C} n_p p$$

where for each point $p \in C$, n_p is an integer and $n_p = 0$ for all but finitely many points.

The degree of D is defined by

$$\deg D = \sum_{p \in C} n_p.$$

For a meromorphic function we write its divisor as (f). Not all divisors come from meromorphic functions. In the first place $\deg(f) = 0$. This comes from the geometric interpretation of f as a map $f : C \to \mathbf{P}_1$. The number of zeros (counted with multiplicities) is the corresponding count of the number of points in $f^{-1}(0)$ which is the degree of f considered as a map to $\mathbf{P}_1$. But this is the same number for $f^{-1}(\infty)$ or any other point.

More importantly, even if the degree of D is zero, there may not be a meromorphic function with the same divisor. On $\mathbf{P}_1$ it is true, because given any points $a, b \in \mathbf{C}$, $(z-a)/(z-b)$ is a meromorphic function with a simple zero at a and a simple pole at b, so for any two points $p, q \in \mathbf{P}_1$, $p - q$ and hence any sum

$$\sum_i (p_i - q_i)$$

is the divisor of a meromorphic function.

However, suppose that on a curve C (or more generally a Riemann surface) we have a meromorphic function f with a simple zero only at p and a simple pole only at q. Then in the geometrical viewpoint $f : C \to \mathbf{P}_1$, $f^{-1}\{0\}$ consists of one point so the degree of the map is 1 and furthermore there can be no ramification points since in the neighbourhood of $z \mapsto z^n$ a nonzero value has n inverse images. From the Riemann-Hurwitz formula $2 - 2g = 2$ and so this feature can only occur when $g = 0$.

A divisor is said to be effective, or positive, if all the n_p are nonnegative. We can add and subtract divisors in the obvious way and so we write

$$D \geq D'$$

if $D - D'$ is effective. So effective means $D \geq 0$. Clearly if $D \geq D'$, then $\deg D \geq \deg D'$.

Definition 21 *A divisor D is said to be a principal divisor if $D = (f)$ for some meromorphic function f.*

Divisors D, D' are said to be linearly equivalent if $D - D'$ is a principal divisor.

We write $D \sim D'$ for two linearly equivalent divisors. Since $\deg(f) = 0$ in the divisor sense, the degree of two linearly equivalent divisors is the same.

We shall be interested in the effective divisors in a given equivalence class. What does this mean? Consider an effective divisor

$$\sum_{i=1}^{k} m_i p_i$$

where $m_i > 0$. An effective divisor $\sum_i n_i q_i$ is equivalent to this if there is a meromorphic function f such that

$$(f) = \sum_i m_i p_i - \sum_i n_i q_i.$$

In other words D is defined by a meromorphic function whose zeros are precisely the p_i with multiplicity m_i. For any divisor we make the following definition:

Definition 22 *Let D be a divisor on the curve C. Denote by $\mathcal{L}(D)$ the set of meromorphic functions f such that $(f) + D \geq 0$ together with the zero function.*

Proposition 24 *(i) $\mathcal{L}(D)$ is a finite-dimensional vector space.*

(ii) If $\deg D < 0$, then $\mathcal{L}(D) = 0$.

(iii) If $D \sim D'$ then $\dim \mathcal{L}(D) = \dim \mathcal{L}(D')$.

(iv) The projective space $P(\mathcal{L}(D))$ is in one-to-one correspondence with the effective divisors equivalent to D.

Proof: (i) Write D as

$$\sum_i m_i p_i - \sum_i n_i q_i$$

where the m_i, n_i are positive, then we are looking at meromorphic functions f which have a pole of order $\leq m_i$ at p_i and no more poles, and have zeros of order $\geq n_i$ at q_i. In other words, functions f whose zeros cancel the q_i and whose poles are cancelled by the p_i. This set is clearly closed under addition and scalar multiplication, and so forms a vector space $\mathcal{L}(D)$.

At each pole p_i, in a local coordinate z, we can write

$$f(z) = \frac{a_{m_i}}{(z - z_i)^{m_i}} + \ldots + \frac{a_1}{(z - z_i)} + h(z)$$

where h is holomorphic and $f \mapsto (a_{m_i}, \ldots, a_1)$ is a linear map from $\mathcal{L}(D)$ to $\mathbf{C}^{m_i}$. The intersection of the kernels of this finite number of maps consists of holomorphic functions which must be constant and hence at most a one-dimensional space, so the vector space $\mathcal{L}(D)$ must be finite-dimensional.

(ii) If $(f) + D \geq 0$ then $0 \leq \deg(f) + \deg D = \deg D$.

(iii) If $D = D' + (g)$, then $f \mapsto fg$ defines an isomorphism from $\mathcal{L}(D)$ to $\mathcal{L}(D')$.

(iv) If $f \in \mathcal{L}(D)$ then by definition $(f) + D$ is effective and linearly equivalent to D. $\square$

The dimension $\ell(D) = \dim \mathcal{L}(D)$ is a very subtle thing in general and there is no simple formula in terms of the genus g and $\deg D$. The Riemann-Roch theorem however relates a particular pair of these numbers.

5.2 Canonical divisors

The derivative of a meromorphic function is not a function. What do we mean by this? In one coordinate we write $f\varphi_U^{-1} = g(z)$ and then the derivative with respect to z is $g'(z)$. But if we change coordinates then

$$f\varphi_{U'}^{-1} = g(h(u))$$

where $h = \varphi_U\varphi_{U'}^{-1}$ and then the derivative with respect to u is $g'(h(u))h'(u)$ and not $g'(h(u))$.

Definition 23 *A meromorphic differential on a Riemann surface is a collection of meromorphic functions f_U on $\varphi_U(U)$ such that on $\varphi_{U'}(U \cap U')$*

$$f_U = f_{U'}(h(u))h'(u)$$

where $h = \varphi_U\varphi_{U'}^{-1}$.

The derivative df of any meromorphic function is a meromorphic differential but not all are of this form.

Remark: 1. The transformation law is more easily described if we write a differential as $f_U(z)dz$, thinking of dz as the derivative of the locally defined coordinate z.

2. The *residue* of a differential at a pole is independent of the local coordinate. This differs from a function, whose *value* is invariant. The simplest way to see the invariance of the residue is to note that the transformation law for dz means that the contour integral

$$\int_\Gamma f dz$$

is well-defined and independent of coordinates. Cauchy's residue theorem then gives the invariance.

Two differentials are defined by local functions f_U, g_U satisfying the conditions in Definition 23. Then

$$\frac{f_U}{g_U} = \frac{f_{U'}}{g_{U'}}$$

on $U \cap U'$ and this defines a meromorphic function – the ratio of two meromorphic differentials is a meromorphic function. This means that the divisors of any two differentials are linearly equivalent, and this is called the *canonical divisor class*.

Examples:

1. The function $z = z_0/z_1$ on $\mathbf{P}^1$ is a meromorphic function with a simple pole at $[1, 0]$, because in the coordinate $\tilde{z} = z_1/z_0$ near $[1, 0]$ it is $1/\tilde{z}$. Its derivative is the differential

$$dz = d(\tilde{z}^{-1}) = -\frac{1}{\tilde{z}^2} d\tilde{z}$$

which has no zeros but has a double pole. A canonical divisor on $\mathbf{P}^1$ therefore has degree -2.

2. The Weierstrass $\wp$-function $\wp(z)$ has a double pole at $z = 0$. Its derivative has a triple pole there and vanishes as we have seen at three points $\omega_1/2, \omega_2/2, (\omega_1 + \omega_2)/2$. The degree of a canonical divisor here is then $3 - 3 = 0$.

Let κ be a canonical divisor, then the vector space $\mathcal{L}(\kappa)$ is isomorphic to the space of holomorphic differentials. Since $\deg \kappa = -2$ for $\mathbf{P}^1$ we see that here there are none. For the torus $\deg \kappa = 0$ which means that any holomorphic differential must have no zeros, and so there is at most a one-dimensional space of them. In fact since

$$d(z + m\omega_1 + n\omega_2) = dz$$

the form dz is such a differential.

More generally:

Proposition 25 *On a nonsingular algebraic curve of genus g, the degree of a canonical divisor is $2g - 2$.*

Proof: We use the map $f : C \to \mathbf{P}_1$ as in the proof of the degree-genus formula defined by $[x, y] \in \mathbf{P}_1$. Let $[a, 1] \in \mathbf{P}_1$ be a point which is not a branch point and consider the meromorphic function

$$g = \frac{y}{x - ay}.$$

(All we have done is composed f with a projective transformation of $\mathbf{P}_1$).

Then this meromorphic function has n simple poles so dg has n double poles. But g' vanishes at the ramification points since by choice none of these lie over ∞, and in the proof of Theorem 23 there are $n(n-1)$ of these. Hence the degree of the divisor of dg is

$$n(n-1) - 2n = 2g - 2$$

from Theorem 23. □

The only differentials we have encountered so far are derivatives of meromorphic functions, which can only have poles of order 2 or more since they are the derivatives of $(z-c)^{-m}$. But on $\mathbf{P}_1$ the differential dz/z has simple poles at 0 and ∞. If we tried to integrate this we would get $\log z$ which is not meromorphic and not even single-valued. So differentials with simple poles do exist, but there is a constraint. In particular:

Proposition 26 *A meromorphic differential cannot have a single simple pole.*

Proof: Suppose for a contradiction that p is the pole of the differential ω. It has non-zero residue and so taking a coordinate neighbourhood of p, and surrounding it with a small contour Γ, we have

$$\int_\Gamma \omega \neq 0.$$

Now triangulate C such that each triangle lies in a coordinate neighbourhod and p lies in the interior of one, Δ_0. By Cauchy's theorem the integral of ω around each triangle $\Delta_i, i \neq 0$ is zero and the integrations along adjacent adges of different triangles cancel (this is like the proof of the Gauss-Bonnet Theorem in the *Geometry of Surfaces* course.) But then the integral around Δ_0 vanishes which is a contradiction. □

Remark: It is clear that this argument can be extended to show that the sum of the residues of a meromorphic differential is always zero.

5.3 Riemann-Roch

This is the theorem we aim to prove about the dimension $\ell(D)$ of $\mathcal{L}(D)$:

Theorem 27 *(Riemann-Roch) Let D be any divisor on a nonsingular projective algebraic curve in $\mathbf{P}_2$ and let κ be a canonical divisor, then*

$$\ell(D) - \ell(\kappa - D) = \deg D + 1 - g.$$

The theorem doesn't tell us what the value of $\ell(D)$ is, which can jump up and down depending on the location of the points, but it tells us the difference of two such numbers. This can be used to great effect. For example, if $\deg D > g-1$ the right hand side is positive so we know that $\ell(D) \geq 0$. If $\deg D > 2g-2$, then $\deg(\kappa - D) < 0$ so $\ell(\kappa - D) = 0$ and then we have an exact formula $\ell(D) = \deg D + 1 - g$.

There is one class of divisors where we can already find lots of elements in $\mathcal{L}(D)$, and we shall need this for the proof of the theorem. Take a line L in $\mathbf{P}_2$ and consider the divisor

$$H = \sum_{p \in L \cap C} I_p(C, L)p.$$

By Bézout's theorem the degree of H is n, the degree of the curve C.

If $ax + by + cz = 0, a'x + b'y + c'z = 0$ are two lines L, L', then

$$f = \frac{ax + by + cz}{a'x + b'y + c'z}$$

is a meromorphic function, so the divisor of the line L is linearly equivalent to the divisor of L'.

Any algebraic curve $Q(x, y, z) = 0$ of degree m defines a divisor in the same way, and

$$\frac{Q(x, y, z)}{(ax + by + cz)^m}$$

is a meromorphic function, so the divisor of Q is in the class of mH. Two such polynomials Q and Q' define the same function on C if and only if they are divisible by P, so we can find easily a large subspace of $\mathcal{L}(mH)$ just by counting polynomials.

Writing

$$Q(x, y, z) = b_0(y, z) + b_1(y, z)x + \ldots + b_m(y, z)x^m$$

the coefficient $b_i(y, z)$ is homogeneous of degree $m-i$ and so has $m-i+1$ coefficients. Thus the vector space of all Q has dimension $1 + 2 + (m + 1) = (m + 1)(m + 2)/2$. The subspace of all $Q = PR$ for R of degree $m - n$ is then of dimension $(m - n + 1)(m - n + 2)/2$ and so

$$\ell(mH) \geq \frac{1}{2}((m + 1)(m + 2) - (m - n + 1)(m - n + 2)) = mn - \frac{1}{2}n(n - 3). \quad (13)$$

The degree of H is the number of points of intersection of C with a line which is n, so $\deg(mH) = mn$. From the degree-genus formula we see that $mn - n(n - 3)/2 = \deg(mH) + 1 - g$ which is the right hand side of the Riemann-Roch formula.

We start the proof of Theorem 27 with a similar-looking proposition:

Proposition 28 *If D is any divisor on C then $\ell(D) - \ell(\kappa - D) \geq \deg D + 1 - g$.*

Proof: We saw above that $\ell(mH) \geq \deg(mH)+1-g$. Moreover if m is large enough $\deg(\kappa - mH) < 0$ so $\ell(\kappa - mH) = 0$. Therefore we already have the inequality for $D = mH$.

In the general case

$$D = \sum_{i=1}^{k} m_i p_i - \sum_{i=1}^{\ell} n_i q_i$$

choose lines $a_i x + b_i y + c_i z = 0$ that pass through the points p_i then with $m = \sum_i m_i$ the divisor of

$$(a_1 x + b_1 y + c_1 z)^{m_1} (a_2 x + b_2 y + c_2 z)^{m_2} \ldots (a_k x + b_k y + c_k z)^{m_k}$$

is of the form

$$\sum_{i=1}^{k} m_i p_i + \sum_{j=1}^{N} r_j = D + x_1 + x_2 + \ldots + x_r.$$

Furthermore, this divisor is linearly equivalent to mH and so

$$\ell(mH) = \ell(D + x_1 + x_2 + \ldots + x_r).$$

By adding more x_i we can assume that m is large enough that $\deg(\kappa - mH) < 0$. We now need the following lemma:

Lemma 29 *For any point p,*

$$0 \leq \ell(D + p) - \ell(\kappa - D - p) - \ell(D) + \ell(\kappa - D) \leq 1$$

Proof: Firstly $f \in \mathcal{L}(D)$ if and only if $(f)+D \geq 0$ which clearly implies $(f)+D+p \geq 0$, so that $\mathcal{L}(D) \subseteq \mathcal{L}(D + p)$ and

$$\ell(D + p) \geq \ell(D).$$

Suppose

$$D = \sum_{i=1}^{k} m_i p_i - \sum_{i=1}^{\ell} n_i q_i.$$

Take $f \in \mathcal{L}(D + p)$. If p is not one of the p_i or q_j then f has at most a simple pole at p. The condition for f to lie in $\mathcal{L}(D)$ is thus a single linear condition, the vanishing

of the coefficient of $(z-a)^{-1}$. If $p = p_1$ say, then in the Laurent expansion around p we have:

$$f(z) = \frac{a_{m_1+1}}{(z-z_1)^{m_1+1}} + \ldots + a_0 + \ldots$$

and here for f to lie in $\mathcal{L}(D)$ is the vanishing of a_{m_1+1}. If $p = q_1$ then f has a zero of order at least $(n_1 - 1)$ at q_1 and to lie in $\mathcal{L}(D)$, must have a zero of order n_1. This is again one linear condition. In all cases we see that

$$\dim \mathcal{L}(D+p) \leq \dim \mathcal{L}(D) + 1.$$

Applying this to D and $\kappa - D - p$, we see that the lemma holds so long as we can eliminate the case

$$\ell(D+p) - \ell(D) = 1 \qquad \ell(\kappa - D) - \ell(\kappa - D - p) = 1.$$

Suppose for a contradiction that this holds. Then there is a meromorphic function f with $(f) + D + p \geq 0$ but $(f) + D \not\geq 0$, so $-p$ is the only negative term in $(f) + (D)$. Similarly there is g such that $(g) + \kappa - D \geq 0$ but $(g) + \kappa - D - p \not\geq 0$ which means that p does not appear in the divisor $(g) + \kappa - D$. Thus in

$$0 \leq (f) + D + p + (g) + \kappa - D = (fg) + \kappa + p$$

the positive element p is not cancelled.

But κ is the divisor of a differential ω and this means that $fg\omega$ is a differential with a single simple pole at p, which is impossible from Proposition 26.

From the lemma, we have

$$\ell(D+x_1+\ldots+x_r) - \ell(\kappa - D - x_1 - \ldots - x_r) \leq \ell(D+x_1+\ldots+x_{r-1}) - \ell(\kappa - D - x_1 - \ldots - x_{r-1}) + 1$$

and repeating we see that

$$\ell(D + x_1 + \ldots + x_r) - \ell(\kappa - D - x_1 - \ldots - x_r) \leq \ell(D) - \ell(\kappa - D) + r$$

or, using (**??**),

$$\ell(mH) - \ell(\kappa - mH) \leq \ell(D) - \ell(\kappa - D) + r$$

or

$$\ell(D) - \ell(\kappa - D) \geq \ell(mH) - \ell(\kappa - mH) - r.$$

So, since we know the inequality for mH,

$$\ell(D) - \ell(\kappa - D) \geq \deg(mH) + 1 - g - r = \deg D + 1 - g$$

which establishes Proposition 28. □

The Riemann-Roch theorem follows directly from this: use the Proposition for D and for $\kappa - D$ then

$$\ell(D) - \ell(K - D) \geq \deg D + 1 - g$$

and

$$\begin{aligned} \ell(\kappa - D) - \ell(D) &\geq \deg(\kappa - D) + 1 - g \\ &= 2g - 2 - \deg D + 1 - g \\ &= -\deg D + g - 1 \end{aligned}$$

so equality holds.

5.4 Applications

The first consequence is another interpretation of the genus g:

Theorem 30 *The vector space of holomorphic differentials on a nonsingular algebraic curve has dimension g, the genus of the curve.*

Proof: This dimension is $\ell(\kappa)$, so take $D = 0$ and use Riemann-Roch:

$$\ell(0) - \ell(\kappa) = 1 - g.$$

But $\mathcal{L}(0)$ consists of holomorphic functions on C which are just the constants and hence one-dimensional, so $\ell(0) = 1$, and so Riemann-Roch gives $\ell(\kappa) = g$. □

We can actually write down these differentials. First consider the affine part of the curve given by $P(x, y, 1) = 0$. Then x is a local coordinate where $\partial P/\partial y \neq 0$ so consider the differential

$$\omega = \frac{dx}{\partial P/\partial y(x, y, 1)}.$$

At first sight this seems to have poles where the denominator vanishes but this is just where the role of x as a local coordinate breaks down. Since the curve is nonsingular, at such points $\partial P/\partial x \neq 0$ and from the chain rule, on the curve $(\partial P/\partial x)dx + (\partial P/\partial y)dy \equiv 0$, so that ω can also be written, using y as a coordinate, as

$$\omega = -\frac{dy}{\partial P/\partial x(x, y, 1)}.$$

This form has no poles and no zeros in the affine part of the curve.

Now look at C near $z = 0$. We have

$$\frac{d(x/z)}{\partial P/\partial y(x/z, y/z, 1)} = \frac{d(x/z)}{\partial P/\partial y(1, y/x, z/x)(x/z)^{n-1}} = \frac{-d(z/x)(x/z)^2}{\partial P/\partial y(1, y/x, z/x)(x/z)^{n-1}}$$

and so

$$\omega = \frac{-z^{n-3}dz}{\partial P/\partial y(1, y, z)}$$

and has a zero of order $n-3$ where $z = 0$.

This tells us that $\kappa \sim (n-3)H$, and so we can obtain a holomorphic differential by writing

$$\frac{Q(x, y, 1)dx}{\partial P/\partial y(x, y, 1)}.$$

for a homogeneous polynomial $Q(x, y, z)$ of degree $n-3$. The dimension of the space of polynomials of this degree is $(n-2)(n-1)/2$ which is g from the degree-genus formula. Riemann-Roch therefore tells us that *every* holomorphic differential is obtained from a polynomial this way.

Example: For a cubic curve $n = 3$ and the construction gives, up to a constant multiple, the unique holomorphic differential. If the cubic is

$$y^2 = 4(x - e_1)(x - e_2)(x - e_3)$$

then this is $dx/2y$, and if $y = \wp'(z)$, $x = \wp(z)$ as in section 4.4, this is the differential $dz/2$.

Example: When the curve C is a quartic, i.e. $n = 4$, then $\kappa \sim H$, so the zeros of a holomorphic differential are the intersections of C with a line.

If $m > (n-3)$ then $\deg(\kappa - mH) < 0$ so Riemann-Roch gives

$$\ell(mH) = mn + 1 - g = mn - n(n-3)/2.$$

This is the lower bound we obtained in (13) by explicitly writing down polynomials so we have a concrete construction also for the divisor class mH.

Another corollary of the theorem provides a link between the complex analysis and the algebra. We have been considering meromorphic functions a great deal and these are defined as analytical or geometrical objects. They are in fact all expressed in terms of polynomials:

Theorem 31 *Any meromorphic function on a nonsingular curve C is expressible as a ratio of homogeneous polynomials of the same degree:*

$$f = \frac{Q(x,y,z)}{R(x,y,z)}.$$

Proof: We showed using Riemann-Roch that $\mathcal{L}(mH)$ for $m \geq n-3$ is generated by polynomials, in other words every meromorphic function f with

$$(f) + mH \geq 0$$

is of the form $Q(x,y,z)/(ax+by+cz)^m$.

If f is any meromorphic function, then taking lines L_i which pass through the poles of f, we have, for some m

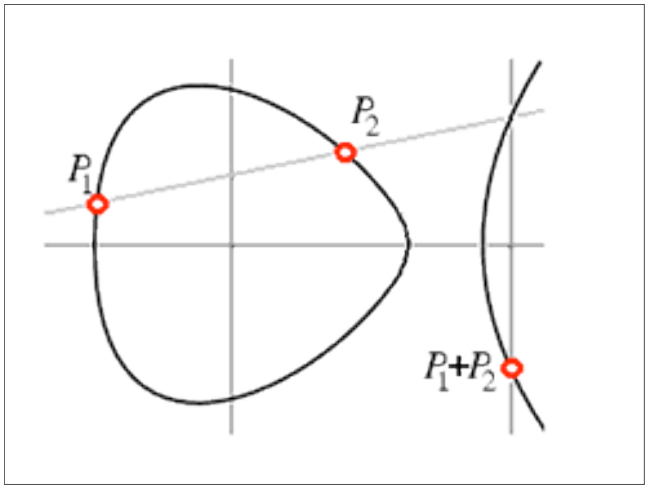
$$(f) + H_1 + \ldots + H_m \geq 0$$

and the same argument shows that f is a rational function. □

5.5 The group law on a cubic

An important property of a cubic curve is that the points on it form an abelian group. It has a very geometrical description:

Theorem 32 *Let C be a nonsingular curve of degree 3 and let e be an inflection point. There is a unique additive group structure on C such that e is the identity element and $p_1+p_2+p_3 = 0$ if and only if p_1, p_2, p_3 are the three points of intersection (counting multiplicities) of C with a line.*



Proof: The addition of divisors is commutative and associative. This is also true of their linear equivalence classes since if $p = p' + (f), q = q' + (g)$ then $p+q = p'+q'+(fg)$. We noted earlier that p is linearly equivalent to q only if $g = 0$, so for any curve C of genus $g > 0$ the equivalence class of p determines p uniquely. We could also use Riemann-Roch: if $D = p$ then since $\deg D = 1 > 0 = \deg \kappa$ we have $\ell(\kappa - D) = 0$ and

$$\ell(D) = 1 + 1 - 1 = 1.$$

For the cubic, with $g = 1$, we take an inflection point e and map $p \mapsto [p - e]$ into the group of equivalence classes of degree zero divisors. From the above, this is injective. Moroever $[e - e] = [0]$ is clearly an identity.

Then $p + q$ maps to $[p + q - 2e]$ and we want to show that this is of the form $[s - e]$. The line $ax + by + cz = 0$ joining p and q (or the tangent at p if $p = q$) meets the degree 3 curve in a third point r by Bézout's theorem. Let $a'x + b'y + c'z = 0$ be the tangent at e, then the divisor of its intersection with C is $3e$ since e is an inflection point.

If $f = (ax + by + cz)/(a'x + b'y + c'z)$, the divisor of f is

$$(f) = p + q + r - 3e$$

which shows that $[p + q - 2e] = [e - r]$.

Now take $q = e$ in this expression, then $[p - e] = [e - p']$ for some p' which we call the inverse of p.

In general then

$$[p + q - 2e] = [e - r] = [r' - e]$$

as required, proving that C is closed under the addition law. □

In Section 4.4 we showed how a torus $\mathbf{C}/\Gamma$ could be mapped isomorphically to a cubic curve by the Weierstrass $\wp$-function. Since $\mathbf{C}/\Gamma$ is clearly an abelian group it seems reasonable to believe that this is the addition law described geometrically above. This will be so if $(1, \wp(u), \wp'(u)), (1, \wp(v), \wp'(v)), (1, \wp(w), \wp'(w))$ are linearly dependent if $w = -(u + v)$, so consider

$$\det \begin{pmatrix} 1 & \wp(u) & \wp'(u) \\ 1 & \wp(v) & \wp'(v) \\ 1 & \wp(-u-v) & \wp'(-u-v) \end{pmatrix}.$$

Fix v and vary u, then this has a pole at $u = 0$ and $\wp(u) = 1/u^2 + a(u)$ where $a(u)$ is holomorphic. So expanding the determinant gives

$$-\frac{1}{u^2}(\wp'(-u-v) - \wp'(v)) - \frac{2}{u^3}(\wp(-u-v) - \wp(v)) + h(u)$$

where $h(u)$ is holomorphic near u. But

$$\wp(-u-v) = \wp(u+v) = \wp(v) + u\wp'(v) + u^2\wp''(v)/2 + u^3\wp'''(v)/6 + \dots$$

so this is

$$-\frac{1}{u^2}(-2\wp'(v) - u\wp''(v) - u^2\wp'''(v)/2) - \frac{2}{u^3}(u\wp'(v) + u^2\wp''(v)/2 + u^3\wp'''(v)/6) + k(u)$$

where $k(u)$ is holomorphic. But the singular terms cancel so the determinant is finite near $u = 0$, and similarly near $u = -v$. But then it is holomorphic everywhere and hence constant. But when $u = v$ it vanishes, and so it is identically zero.

Remark: The inflection points are the points p such that the divisor of a line is $3p$, or in the group law $3[p - e] = 0$. Given that the group law is addition in $\mathbf{C}/\Gamma$ these are the nine points of order 3 in the group.

.