Bezout's theorem

1 Bezout's theorem

Let C and D be two plane curves, described by equations $f(X,Y)=0$ and $g(X,Y)=0$, where f and g are nonzero polynomials of degree m and n, respectively. Bezout's theorem says that if all is well, then C and D meet in precisely mn points.

1.1 Too many points in common

What can go wrong? If $C=D$ then both curves have all points in common, probably infinitely many (depending on the field).

Also, if the equation factors, then the curve is the union of several components. For example, if C is given by the equation $XY=0$, then it is the union of the two lines given by $X=0$ and $Y=0$. And if D is given by the equation $Y^3-X^3Y=0$, then it is the union of the line $Y=0$ and the cubic curve $Y^2=X^3$. Now C and D have the line $Y=0$ in common.

But this is the only way there can be too many points in common. So the first statement is: if C and D do not have a common component, then they have at most mn points in common.

1.2 Too few points in common

There are several ways there can be 'missing' points of intersection. Missing points can have coordinates in an extension field, they can lie at infinity, and they can coincide with other common points.

For example, a straight line and a circle may meet in two points, and all is well. Or in zero points, which means that the quadratic equation one has to solve to find the points of intersection has a discriminant that is not a square in the field, and the two points of intersection have coordinates in a quadratic extension field. Finally, the straight line can be tangent to the circle, and we must count the point of intersection twice.

And, for example, two parallel lines have a common point at infinity.

So now Bezout's theorem becomes: *If C and D do not have a common component, then they have at most mn points in common. If the field is algebraically closed, and we also count points at infinity, and we count intersection points with proper multiplicity, then there are precisely mn common points.*

1.3 The inequality

Let $R=k[X,Y]$ be the ring of polynomials in the two variables X,Y with coefficients in the field k.

Proposition 1.1 *Let $f, g \in R$ be nonzero polynomials of degrees m, n, respectively. Let C and D be two plane curves, described by the equations $f(X, Y) = 0$ and $g(X, Y) = 0$. If f and g do not have a common factor, then $|C \cap D| \leq \dim_k R/(f, g) \leq mn$.*

Proof:

(1) *Given distinct points P_i $(1 \leq i \leq t)$, there are polynomials $h_i \in R$ $(1 \leq i \leq t)$ such that $h_i(P_i) \neq 0$ and $h_i(P_j) = 0$ for all i, j, $i \neq j$.*

(Indeed, if $P_i = (x_i, y_i)$, put $h_i(X, Y) = \prod_{x_j \neq x_i}(X - x_j) . \prod_{y_j \neq y_i}(Y - y_j)$.)

(2) $|C \cap D| \leq \dim_k R/(f, g)$.

(Indeed, if C and D have common points P_i, then make polynomials as in (1). If $\sum c_i h_i = uf + vg$ with $u, v \in R$, then substitute the P_i to find $c_i = 0$. This means that the images $h_i + (f, g)$ of the h_i in $R/(f, g)$ are linearly independent.)

(3) *Let R_d be the k-vectorspace of polynomials $p(X, Y)$ of total degree at most d. If $d \geq 0$ then $s(d) := \dim_k R_d = 1 + \cdots + (d + 1) = \frac{1}{2}(d + 1)(d + 2)$.*

(4) *We have $\dim_k R_d/(f, g) \leq mn$ for all d.*

(Indeed, consider the sequence of maps

$$R_{d-m} \times R_{d-n} \xrightarrow{\alpha} R_d \xrightarrow{\pi} R_d/(f, g) \rightarrow 0$$

where α is the map $\alpha(u, v) = uf + vg$ and π is the quotient map. Since f and g do not have a common factor, the kernel of α consists of the pairs $(wg, -wf)$ with $w \in R_{d-m-n}$ and hence has dimension $s(d - m - n)$ for $d \geq m + n$. It follows that the image of α has dimension $s(d - m) + s(d - n) - s(d - m - n)$. Since π is surjective, and $\pi\alpha = 0$, we find $\dim_k R_d/(f, g) \leq s(d) - s(d - m) - s(d - n) + s(d - m - n) = mn$.)

(5) *We have $\dim_k R/(f, g) \leq mn$.*

(Indeed, if we can find more than mn linearly independent elements in $R/(f, g)$, then for sufficiently large d they will be in $R_d/(f, g)$, contradicting (4).)

□

1.4 Points at infinity

1.4.1 Affine and projective space

Extend affine space to projective space by adding points at infinity, as follows. A point in n-dimensional affine space has n coordinates $(x_1, ..., x_n)$ in the underlying field. A point in n-dimensional projective space has $n + 1$ coordinates $(x_1, ..., x_{n+1})$, not all zero, where only the ratio is significant: if $a \neq 0$ then $(x_1, ..., x_{n+1}) = (ax_1, ..., ax_{n+1})$.

1.4.2 Equations

An equation like $Y = X^2$ is meaningful in an affine space, but not in a projective space, since whether equality holds must not change when all cordinates of a point are multiplied by the same nonzero constant. Thus, for a projective space one needs *homogeneous* equations, equations such that all terms have the same degree, as in $YZ = X^2$.

1.4.3 Going back and forth

One can embed n-dimensional affine space into n-dimensional projective space by $(x_1, ..., x_n) \mapsto (x_1, ..., x_n, 1)$. Conversely, if $x_{n+1} \neq 0$ then one can scale coordinates to make $x_{n+1} = 1$, omit this coordinate and find a copy of affine space. The projective points outside this copy, the 'points at infinity', have $x_{n+1} = 0$.

The above describes projective n-space as affine n-space together with points at infinity. That might give the false impression that projective space has two types of points. A homogeneous description identifies the projective point $(x_1, ..., x_{n+1})$ with the line in affine $(n+1)$-space that passes through the origin and the point $(x_1, ..., x_{n+1})$. The above identification now identifies a line with its point of intersection with the hyperplane $x_{n+1} = 1$, and the points at infinity are the lines through the origin that are parallel to that hyperplane.

One goes from affine equation to projective (homogeneous) equation by making all degrees equal (to the maximum degree) by inserting factors x_{n+1}. For example, the homogeneous version of the cubic equation $Y^2 = X^3 - 1$ is $Y^2Z = X^3 - Z^3$. And one goes back to the affine equation by substituting $x_{n+1} = 1$.

Example With affine coordinates (X, Y), consider the line $X = 0$ and the parabola $Y = X^2$. In projective coordinates (X, Y, Z), these equations become $X = 0$ and $YZ = X^2$. The two common points are $(0, 1, 0)$ and $(0, 0, 1)$. The former is the point at infinity of the Y-axis, the latter is the origin $(0, 0)$.

1.4.4 No common points at infinity

Proposition 1.2 *Let k be a field, and let $F, G \in k[X, Y, Z]$ be homogeneous polynomials of degrees m, n, respectively. Put $f = F(X, Y, 1)$ and $f^* = F(X, Y, 0)$ and define g, g^* similarly. If f and g do not have a common factor, and f^* and g^* do not have a common factor, then $\dim_k R/(f, g) = mn$.*

Proof: For a polynomial p, let p^* be the sum of its terms of highest degree. Continue the proof of Proposition 1.1. We want to show that $\dim_k R_d/(f, g) = mn$ for large d, and the argument of (4) will yield that, if the kernel of π is the image of α. Suppose $\pi(h) = 0$ for some $h \in R_d$. Then $h = uf + vg$ for certain $u, v \in R$, and we take u, v of smallest degree. If u has degree larger than $d - m$, then the terms of highest degree cancel, so $u^*f^* + v^*g^* = 0$. Since f^* and g^* do not have a common factor, there is a $w \in R$ with $u^* = wg^*$ and $v^* = -wf^*$. Now $h = (u - wg)f + (v + wf)g$ where $u - wg$ and $v + wf$ have smaller degrees than u and v, contradiction. Hence u has degree at most $d - m$, and similarly v has degree at most $d - n$, and h lies in the image of α. □

Remark The condition that f and g do not have a common factor is equivalent to asking that F and G do not have a common factor. The points at infinity of the curve $f = 0$ are the points $(a, b, 0)$ with $f^*(a, b) = 0$, that is, such that f^* has a factor $aY - bX$. The requirement that f^* and g^* do not have a common factor is equivalent to asking that the curves $f = 0$ and $g = 0$ do not have common points at infinity with coordinates in the algebraic closure $\bar{k}$ of k.

1.5 Intersection multiplicity

Let C and D be two plane curves defined by $f(X,Y)=0$ and $g(X,Y)=0$ and let P be a point. We want to define the intersection multiplicity $I_P(f,g)$ of C and D at the point P. It should be a nonnegative integer, or ∞ in case C and D have a common component that passes through P.

First an operational definition, a series of rules that suffice to compute $I_P(f,g)$. We have $I_P(f,g)=I_P(g,f)$, and $I_P(f,g+fh)=I_P(f,g)$, and $I_P(f,gh)=I_P(f,g)+I_P(f,h)$ and $I_P(f,g)=0$ if P is not a common point of C and D, and $I_P(f,g)=1$ if C and D are nonsingular at P with distinct tangents, and $I_P(f,g)=\infty$ iff f and g have a common factor.

Example Consider the two circles $X^2+Y^2=1$ and $X^2+Y^2=2$. Clearly, any common points must lie at infinity. The homogeneous equations are $X^2+Y^2-Z^2=0$ and $X^2+Y^2-2Z^2=0$, and the common points are the two points $(\pm i,1,0)$. Now let $P=(i,1,0)$ and consider the intersection multiplicity at P. We have $I_P(X^2+Y^2-Z^2,X^2+Y^2-2Z^2)=I_P(X^2+Y^2-Z^2,Z^2)=I_P(X^2+Y^2,Z^2)=2I_P(X^2+Y^2,Z)=2I_P(X+iY,Z)+2I_P(X-iY,Z)=0+2=2$. So, the four common points are the two points $(\pm i,1,0)$, each counted twice.

Example Consider the two curves $Y=X^3$ and $Y=X^5$. The homogeneous equations are $YZ^2=X^3$ and $YZ^4=X^5$, and the common points are the points $(0,0,1)$, $(1,1,1)$, $(-1,-1,1)$, $(0,1,0)$. Since $(1,1,1)$ and $(-1,-1,1)$ are ordinary points on the curves, and the curves have different tangents at each of these points, the intersection multiplicity at $(1,1,1)$ and $(-1,-1,1)$ is 1. Let $P=(0,0)$. Then $I_P(Y-X^3,Y-X^5)=I_P(Y-X^3,X^3-X^5)=I_P(Y-X^3,X^3)+I_P(Y-X^3,1-X^2)=3I_P(Y,X)+0=3$, so that the origin is a point with intersection multiplicity 3. Let $Q=(0,1,0)$ and make this the origin by choosing affine coordinates $(X/Y,Z/Y)$, that is, by putting $Y=1$. Then $I_Q(X^3-Z^2,X^5-Z^4)=I_Q(X^3-Z^2,X^5-X^3Z^2)=I_Q(X^3-Z^2,X^3)+I_Q(X^3-Z^2,X^2-Z^2)=I_Q(Z^2,X^3)+I_Q(X^3-X^2,X^2-Z^2)=6I_Q(Z,X)+4I_Q(X,Z)+0=10$. So, the fifteen common points are the two points $(1,1)$ and $(-1,-1)$ each counted once, the point $(0,0)$ counted three times, and the point at infinity of the Y-axis $(0,1,0)$, counted ten times.

Algorithm Note that the operational definition will always compute some answer. We may assume that f and g do not have a common factor and that $f(P)=g(P)=0$. If $P=(x,y)$ then consider the polynomials $f^*(X):=f(X,y)$ and $g^*(X):=g(X,y)$. We may suppose f^* has degree not larger than that of g^*.

If f^* is the zero polynomial, then f has a factor $(Y-y)$, and $I_P(f,g)=I_P(Y-y,g^*)+I_P(f_0,g)$, where $f=(Y-y)f_0$. Since f and g do not have a common factor, g^* is not the zero polynomial, and $g^*(X)=(X-x)^i g_2(X)$ with $i\geq 1$ and $g_2(x)\neq 0$, and now $I_P(Y-y,g^*)=i$. So, finding $I_P(f,g)$ has been reduced to finding $I_P(f_0,g)$.

Since $f\neq 0$, we arrive after finitely many steps in the situation where $f^*\neq 0$. Let f^* have leading term ax^d, and let g^* have leading term bX^e. Put $g_0=g-\frac{b}{a}(X-x)^{e-d}f$. Now $I_P(f,g)=I_P(f,g_0)$ and g_0^* has smaller degree than g^* so by induction we are done.

Induction on what? The degrees of f^* and g^* go down until one of them is zero, then we divide f or g by $(Y-y)$, and afterwards the degree of f^* or g^* may be very large again. But we have the promise that $I_P(f,g)$ is going to be finite, and each time that $f^* = 0$ or $g^* = 0$ we get a contribution of at least 1, so this can happen only finitely many times, and the algorithm terminates.

1.5.1 The local ring at P

A *local ring* is a ring with a unique maximal ideal.

Given a field k and a point $P \in k^2$, let $\mathcal{O}_P$ be the ring of rational functions $\frac{u}{v}$ with $u, v \in R$ and $v(P) \neq 0$. This ring has a unique maximal ideal $M_P = \{\frac{u}{v} \in \mathcal{O}_P \mid u(P) = 0\}$ and is called the *local ring at* P.

1.5.2 Definition of the intersection multiplicity

Let C and D be curves in the plane given by equations $f(X,Y) = 0$ and $g(X,Y) = 0$. Let P be a point. Let $(f,g)_P$ be the ideal $\mathcal{O}_P f + \mathcal{O}_P g$ in $\mathcal{O}_P$ generated by f and g.

Definition $I_P(C,D) = I_P(f,g) = \dim_k \mathcal{O}_P/(f,g)_P$.

Proposition 1.3 *If f, g do not have a common factor, then $\mathcal{O}_P = R + (f,g)_P$ (that is, elements of $\mathcal{O}_P$ have polynomial representatives), and we have $I_P(f,g) = \dim_k \mathcal{O}_P/(f,g)_P \leq \dim_k R/(f,g)$.*

Proof: Given finitely many elements of $\mathcal{O}_P$, we can write them with common denominator. If the images of $\frac{u_1}{v}, ..., \frac{u_t}{v}$ are linearly independent in $\mathcal{O}_P/(f,g)_P$, then $u_1, ..., u_t$ are linearly independent in $R/(f,g)$ since $\frac{1}{v} \in \mathcal{O}_P$. This proves the statement about dimensions.

Since f, g do not have a common factor, we have $\dim_k R/(f,g) \leq mn$, so $\dim_k \mathcal{O}_P/(f,g)_P$ is finite. If $\frac{u_1}{v}, ..., \frac{u_t}{v}$ is a basis of $\mathcal{O}_P/(f,g)_P$, then (since $v, \frac{1}{v} \in \mathcal{O}_P$ so multiplication by v is invertible) also $u_1, ..., u_t$ is a basis. □

Example Let $f(X,Y) = Y$ and $g(X,Y) = Y - X^3$. The intersection multiplicity of the cubic $Y = X^3$ and the line $Y = 0$ at $P = (0,0)$ should be 3. The quotient ring $R/(f,g)$ is a vector space over k, and the images of 1, X, X^2 form a basis, so $\dim_k R/(f,g) = 3$ and also $\dim_k \mathcal{O}_P/(f,g)_P = 3$.

Example Let $f(X,Y) = Y^2 - X^3$ and $g(X,Y) = Y^3 - X^4$. Then $I_P(f,g) = 8$ for $P = (0,0)$. A Gröbner basis for (f,g) is given by $\{X^3 - Y^2, XY^2 - Y^3, Y^5 - Y^4\}$ so that $R/(f,g)$ has basis with representatives X^2Y, X^2, XY, X, Y^4, Y^3, Y^2, Y, 1, and $\dim_k R/(f,g) = 9$. But $Y - 1$ is nonzero in P, so $(f,g)_P$ also contains $(Y^5 - Y^4)/(Y - 1) = Y^4$, and $\dim_k \mathcal{O}_P/(f,g)_P = 8$.

In these examples it was clear that $\dim_k \mathcal{O}_P/(f,g)_P$ had at most the given value. That it has precisely the claimed value will follow if we show that $I_P(f,g)$ defined in this way satisfies the rules given earlier.

Proposition 1.4 *The algorithm given earlier computes $I_P(f,g)$ as defined above. The rules given earlier are valid.*

Proof: The rules $I_P(f,g) = I_P(g,f)$ and $I_P(f, g + fh) = I_P(f,g)$ are obvious, since the ideal $(f,g)_P$ does not change.

If f and g have a common factor h, $\dim_k \mathcal{O}_P/(f,g)_P \geq \dim_k \mathcal{O}_P/(h)_P = \infty$. Conversely, if f and g do not have a common factor, then $\dim_k \mathcal{O}_P/(f,g)_P \leq \dim_k R/(f,g) \leq mn < \infty$ if f and g have degrees m and n, respectively.

For the rule $I_P(f,gh) = I_P(f,g) + I_P(f,h)$ consider the sequence

$$0 \to \mathcal{O}_P/(f,h)_P \overset{*g}{\to} \mathcal{O}_P/(f,gh)_P \to \mathcal{O}_P/(f,g)_P \to 0$$

where the second arrow is multiplication by g, and the third is the quotient map. If we show that this sequence is exact, then by taking dimensions our rule follows. For exactness, the only nontrivial part is showing that $*g$ is injective. If for some $z \in \mathcal{O}_P$ we have $zg \in (f,gh)_P$, say $zg = uf + vgh$ where $u, v \in \mathcal{O}_P$, then multiply by the denominators to get the relation $\bar{z}g = \bar{u}f + \bar{v}gh$ with $\bar{z}, \bar{u}, \bar{v} \in R$. We may assume that f and g do not have a common factor. Now $g|\bar{u}$ and $\bar{z} = (\bar{u}/g)f + \bar{v}h \in (f,h)$ and we had $z \in (f,h)_P$ before eliminating denominators.

If $f(P) \neq 0$, then $\frac{1}{f} \in \mathcal{O}_P$, so $1 \in (f,g)_P$ and $\mathcal{O}_P/(f,g)_P = (0)$ and $I_P(f,g) = \dim_k(0) = 0$ as desired.

The final item needed in the algorithm was $I_P(X-x, Y-y) = 1$ for $P = (x,y)$. W.l.o.g., take $P = (0,0)$. Now $R/(X,Y) \simeq k$ and $\mathcal{O}_P/(X,Y)_P \simeq k$, so $\dim_k \mathcal{O}_P/(X,Y)_P = 1$ as desired.

That proves the claim about the algorithm. Remains the last rule: $I_P(f,g) = 1$ if C and D are nonsingular at P with distinct tangents. W.l.o.g., take $P = (0,0)$. If we follow the given algorithm and f and g have nonproportional linear parts, this remains true upon replacing g by $g - cX^d f$, and after finitely many steps we reach $I_P(X,Y) = 1$. □

1.5.3 Inequality with multiplicities

As before, let f, g be polynomials of degrees m, n, respectively, and assume that f and g do not have a common factor. This means that the curves C and D defined by $f = 0$ and $g = 0$, respectively, have only finitely many points (namely at most mn) in common.

Proposition 1.5 *We have*

$$\sum_P I_P(f,g) = \sum_P \dim_k \mathcal{O}_P/(f,g)_P \leq \dim_k R/(f,g) \leq mn$$

where the sums run over $P \in C \cap D$.

Proof: We show that the natural map $R \to \prod_P \mathcal{O}_P/(f,g)_P$ (that sends $h \in R$ to the element with P-coordinate $h + (f,g)_P$) is surjective. Then the statement of the proposition follows by taking dimensions (since (f,g) is in the kernel of this map).

In order to show surjectivity it suffices to find for any P and arbitrary $z \in \mathcal{O}_P$ an element $h \in R$ that maps to the element $(0, ..., 0, z + (f,g)_P, 0, ..., 0)$ with all coordinates 0 but with P-coordinate $z + (f,g)_P$.

Start by finding polynomials h_P with $h_P(P) = 1$ and $h_P(Q) = 0$ when P, Q are distinct points in $C \cap D$, as in (1) of the proof of Proposition 1.1. If there is a natural number N such that $h_P^N \in (f,g)_Q$ for $Q \neq P$, then pick a polynomial representative p of $zh_P^{-N} \in \mathcal{O}_P/(f,g)_P$ (as we can by Proposition 1.3) and put $h = ph_P^N$. This h satisfies the requirements.

Remains to show the existence of N. It suffices to show: *Let p be a polynomial with $p(Q) = 0$. Let $N \geq d := \dim_k \mathcal{O}_Q/(f,g)_Q$. Then $p^N \in (f,g)_Q$.* Indeed, let $J_i := p^i\mathcal{O}_Q + (f,g)_Q$. The sequence $(J_i)_{i\geq 0}$ of ideals is decreasing, but can have at most $d+1$ different members, so there is an i with $0 \leq i \leq d$ such that $J_i = J_{i+1}$. This means that $p^i = p^{i+1}u + v$ with $u \in \mathcal{O}_Q$ and $v \in (f,g)_Q$. Since $\frac{1}{1-pu} \in \mathcal{O}_Q$ it follows that $p^i = \frac{v}{1-pu} \in (f,g)_Q$, as desired. □

1.5.4 Equality

Proposition 1.6 *Assume that the field k is algebraically closed. Then*

$$\sum_P \dim_k \mathcal{O}_P/(f,g)_P = \dim_k R/(f,g).$$

Proof: We have to show that (f,g) is the full kernel of the map $\pi : R \to \prod_P \mathcal{O}_P/(f,g)_P$. Pick h in this kernel. Consider $L := \{p \in R \mid ph \in (f,g)\}$. This is an ideal in R. If $(x,y) \in V(L)$ then $P := (x,y) \in C \cap D$ since $f,g \in L$. Since $\pi(h)_P = 0$ we have $h = uf + vg$ for certain $u,v \in \mathcal{O}_P$. Write u,v with common denominator p, then $p \in L$ and $p(P) \neq 0$, contradiction. Hence $V(L) = \emptyset$. Since k is algebraically closed we can apply the Nullstellensatz and conclude that $1 \in L$, that is, $h \in (f,g)$. □

1.6 Conclusion

We did all the work required. Assume that k is algebraically closed. The curves C and D have only finitely many points (affine or at infinity) in common, and there is a line that misses all these points (since k is infinite). Choose such a line as line at infinity to find that C and D have precisely mn points in common, counting multiplicities.

(A detail to check: is the definition of intersection multiplicity invariant for change of coordinates? But it is.)