Bézout's Theorem: A taste of algebraic geometry

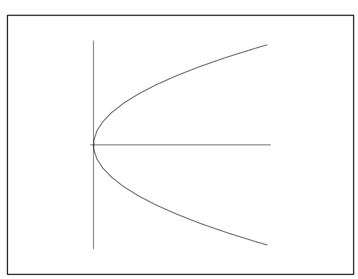
Stephanie Fitchett
Florida Atlantic University Honors College
sfitchet@fau.edu

Abstract: Algebraic geometry is the study of zero sets of polynomials, and can be seen as a merging of ideas from high school algebra and geometry. One of the "Great Theorems" in algebraic geometry is Bézout's Theorem, which explains the intersections of polynomial curves in the (projective) plane. Bézout's Theorem will be illustrated through several examples, followed by a brief discussion of how the tools of modern algebra are used to make intuitive geometric ideas precise.

Introduction

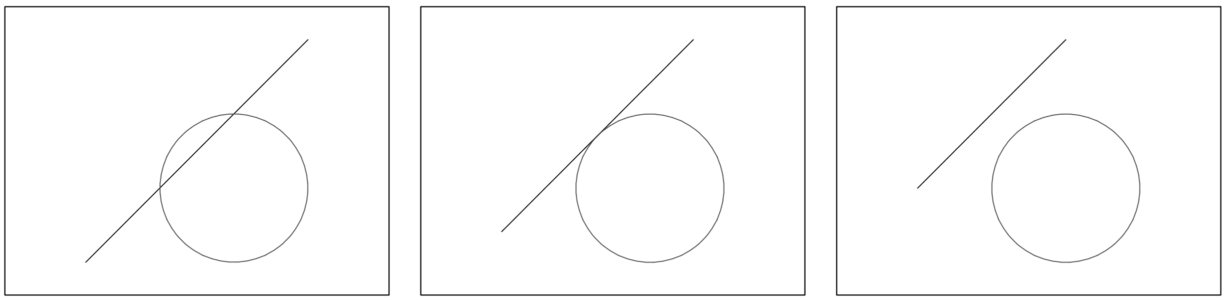
Several recent MAA meetings (winter and summer) have included special sessions on "Great Theorems in Mathematics." The talks in these sessions have been expository, and the presenters have shared their favorite mathematical theorems, which may have beautiful statements, or intriguing proofs, or surprising implications, or some combination of the above. The sessions have been hugely successful; the talks allow those in the audience to appreciate beautiful results in a variety of fields that they may not be intimately familiar with. The discussion that follows is in the spirit of a "Great Theorem" talk: I will not be presenting any new mathematics, but I want to share with you one of the most fundamental and amazing theorems in algebraic geometry. What is wonderful about Bézout's Theorem is not just its statement, but the search for the right hypotheses — those that make the statement of the theorem clean and simple — and the surprising fact that although the statement of the theorem seems entirely geometric, its proof is entirely algebraic. The interplay between geometric intuition and formal algebraic proofs was one of the factors that influenced my decision to study algebraic geometry, and I hope you will enjoy seeing an example of this interplay, whether it is familiar or not.

Algebraic geometry is concerned with the zero sets of polynomials, a topic with which we are all familiar from high school. For instance, the zero set of the polynomial $f(x, y) = x - y^2$ is the curve defined by $f(x, y) = 0$, namely the parabola shown below.



$x = y^2$

Bézout's Theorem is concerned with the intersections of such zero sets. For instance, the pictures below illustrate some different types of intersections between a circle and a line in the plane.



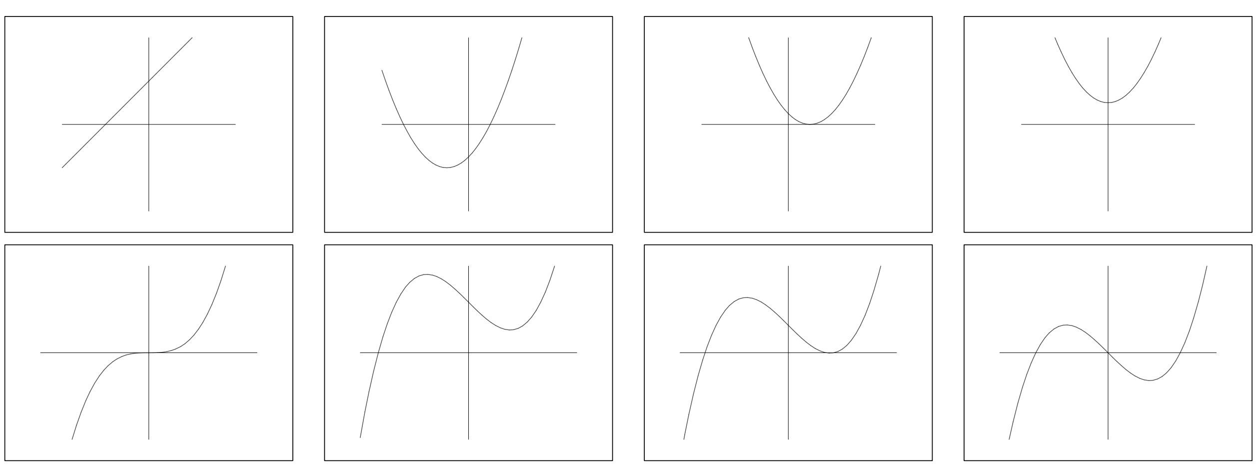
Incidentally, while people study zero sets of polynomials (often called *algebraic varieties*) in arbitrarily large dimensions, and there are analogs of Bézout's Theorem in higher dimensions, we will limit our current discussion to algebraic curves (in a plane) and their intersections, i.e., one-dimensional subvarieties of the plane and their intersections.

The Pre-cursor of Bézout's Theorem: High School Algebra

Let's begin by recalling some basic facts from high school algebra (facts which, while basic to believe, are admittedly not so basic to prove). We know that if $f(x)$ is a non-zero polynomial of degree n, then it has at most n roots.

$$\#(\text{roots of } f) \leq \deg f.$$

For instance, linear polynomials with non-zero slope always have exactly one root, as illustrated below. Quadratic polynomials may have two roots, or one, or none, as illustrated below. Cubic polynomials always have at least one root, but can have no more than three.

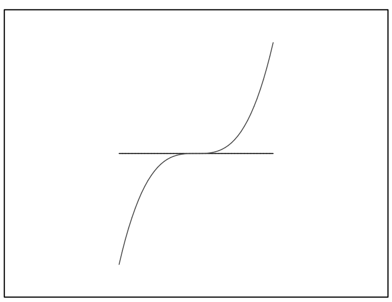


(In fact, we can make an educated guess about the degree of a polynomial, given its graph, based partly on what we know about roots.)

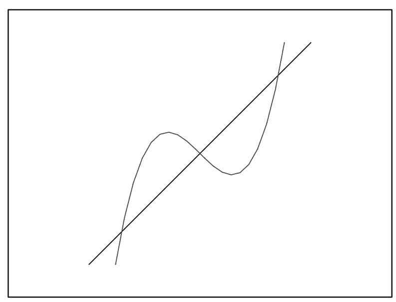
Of course, our pictures (and our counting, thus far) only capture real roots. If we allow non-real roots, and count roots with the appropriate multiplicities (note that for polynomials of low degree we can "see" the multiplicity of a real root r in the way the graph touches or crosses the x-axis at r), we see that

$$\#(\text{roots of } f) = \deg f.$$

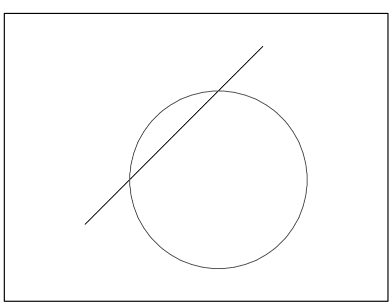
We naturally think of the (real) roots of a polynomial $f(x)$ as the points where the graph of $y = f(x)$ crosses the x-axis. Remembering that Bézout's Theorem is about the *intersections* of curves (and that this discussion is supposed to be leading up to Bézout's Theorem!), let's rephrase the last statement in terms of intersections: The roots of $f(x)$ correspond to the points at which the zero set of the polynomial $y - f(x)$ and the zero set of the polynomial y intersect. This leads naturally to the question, "How do we count the number of points intersection common to *any* two curves?" The accompanying figures illustrate intersections between several pairs of plane curves.



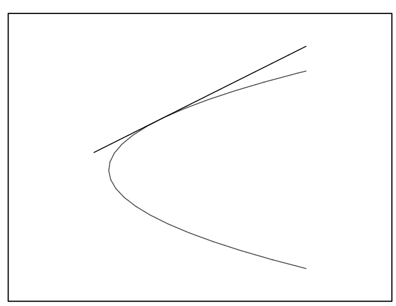
$y = x^3$
$y = 0$



$y = x^3 - x$
$y = x$



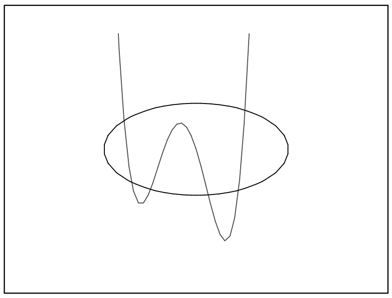
$x^2 + y^2 = 1$
$y = x + 1$



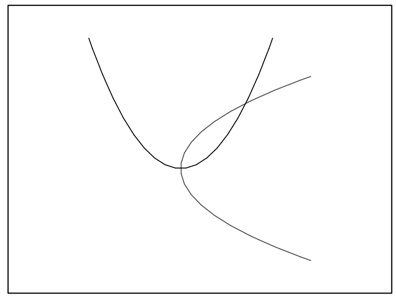
$x = y^2 - 1$
$y = 0.5x + 1$

In the first example, we see a polynomial of degree 3 (namely $y = x^3$) that has just one point in common with the linear polynomial $y = 0$. From our earlier discussion, we might be inclined to count this point with multiplicity 3. Moving down the first column, we see that the unit circle (a degree 2 polynomial) has 2 points in common with the line $y = x + 1$. In the second column, a degree 3 polynomial ($y = x^3 - x$) has three points in common with the line $y = x$, and a quadratic polynomial ($x = y^2 - 1$) has one point in common with the line $y = 0.5x + 1$.

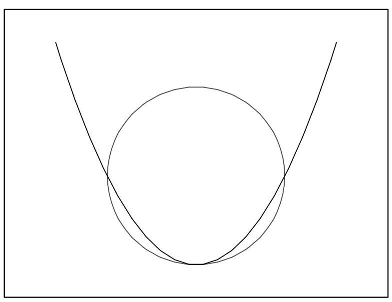
At this point, there does seem to be a pattern: a polynomial of degree n appears to have at most n points in common with a line (a polynomial of degree 1). The next group of examples show intersections between two non-linear curves.



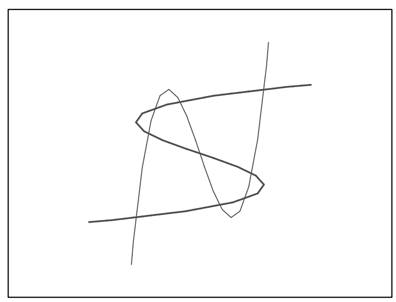
$x^2 + 4y^2 = 4$
$y = 3x(x + 1.5)(x + 7)(x - 1)$



$x = y^2$
$y = x^2$



$x^2 + y^2 = 1$
$y = x^2 - 1$



$y = 3(x^3 - x)$
$x = 3(y^3 - y)$

The ellipse and the quartic illustrate a polynomial of degree 2 and a polynomial of degree 4 that have 6 points in common. [What would be the largest number of possible intersections between an ellipse and the fourth degree polynomial in the example?] The other examples show two quadratic polynomials (a circle and a parabola) which appear to intersect 3 times, two parabolas that intersect twice, and two cubics that intersect 9 times.

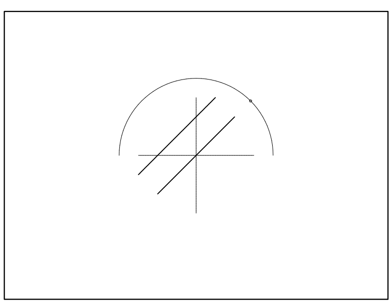
Counting points of intersection in our examples suggests that, for two plane curves C and D, defined by polynomials $f(x, y)$ and $g(x, y)$, respectively,

$$\#(\text{points in } C \cap D) \leq (\deg f)(\deg g).$$

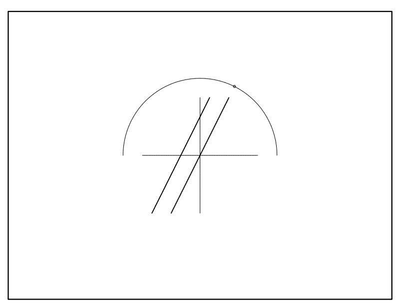
Based on our experience with the fundamental theorem of algebra, we would like to replace the inequality with an equality, and in fact, this is exactly what Bézout's Theorem claims, but we need to find the right hypotheses. In our earlier discussion, we could replace the inequality with an equality provided we allowed non-real roots (points of intersection with the line $y = 0$), and counted roots (points of intersection) with multiplicity. While these extra conditions *do* give equality in all eight examples above, unfortunately, even with these provisions, we cannot replace the inequality with an equality in the case of the intersection of *any* two (algebraic) plane curves. To see why, consider the case of two parallel lines. No matter how carefully we count intersections, two parallel lines simply do not intersect. So to get an equality in our equation, we need stronger assumptions—assumptions which, at the very least, force two parallel lines to "intersect."

The Projective Plane and Homogenization

What would happen if we simply dictated the minimum assumption that is clearly necessary, "Any pair of distinct lines must intersect exactly once"? This is the point of view of projective geometry: we will add "points at infinity" to the regular (affine) plane until any two distinct lines intersect exactly once. We will need to add one point at infinity for parallel lines. Think of adding a point at infinity where $y = x$ and $y = x + 1$ will eventually meet up. Now what about the lines $y = 2x$ and $y = 2x+1$? They will need a point at infinity as well, so we ask, "Can it be the same point that we already added?" Of course the answer must be no, for if $y = 2x$ shared a point at infinity with $y = x$, then the lines $y = 2x$ and $y = x$ would intersect *twice*: once at the origin, and once at a point at infinity. But we surely would not want two lines to intersect *twice*, so the second pair of parallel lines must need their own point at infinity. Following this argument to its logical conclusion, we see that we need exactly one point at infinity for each possible slope of a line. In the pictures below, the semi-circles represent "points at infinity," and the point at infinity where the parallel lines intersect is shown.



$y = x + 1$
$y = x$

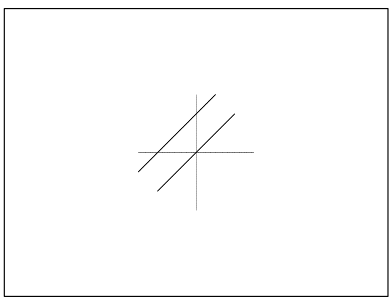


$y = 2x + 1$
$y = 2x$

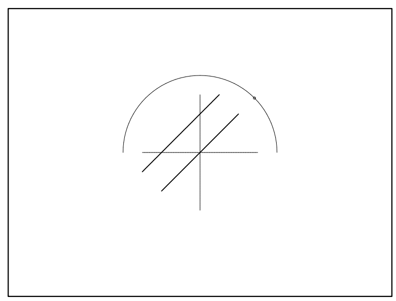
One way to specify coordinates in our new *projective* plane, is as follows. For points (x, y) in the regular plane, specify the same point in the extended plane by $[x, y, 1]$. For a point at infinity which is contained in lines of slope y/x, specify the point by $[x, y, 0]$. Note that this gives exactly one point for each point in the regular plane, plus exactly one point for each possible slope of a line (vertical lines contain the point $[0, 1, 0]$). Also note that because there are many ways to express a slope y/x with different values of y and x (2, for example, can be expressed $2/1$ or $4/2$ or $-10/-5$, among many others), each of these different expressions must correspond to the same point. More generally, points in the projective plane have three coordinates $[a, b, c]$, not all of which are zero, and two characterizations $[a, b, c]$ and $[a', b', c']$ represent the same point if $a' = ka$, $b' = kb$ and $c' = kb$ for a non-zero constant k.

Of course, if our points have three coordinates, our equations will need three variables. We accomplish this by the process of *homogenization*. To homogenize a polynomial equation of degree d, multiply every term with degree less than d by exactly the appropriate power of Z to make the term have degree d. We generally use capital letters to denote variables in the homogenized equation and lower case letters for variables in the dehomogenized equation. Some examples will help clarify the process.

Example 1. The system $y = x + 1$ and $y = x$ is homogenized to $Y = X + Z$ and $Y = X$, which reduces to $Y = X$ and $Z = 0$, so there is a single point of intersection at infinity, namely $[1, 1, 0]$ (which could also be expressed $[k, k, 0]$ for any non-zero constant k). This point is seen as the point at which a line with slope 1 would intersect the 'line' at infinity ($z = 0$, represented by a semi-circle in the picture).

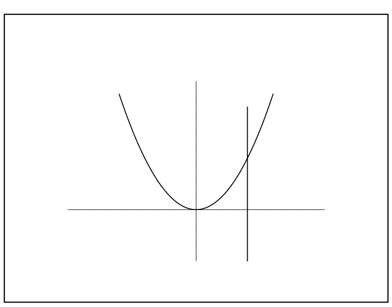


$y = x + 1$
$y = x$

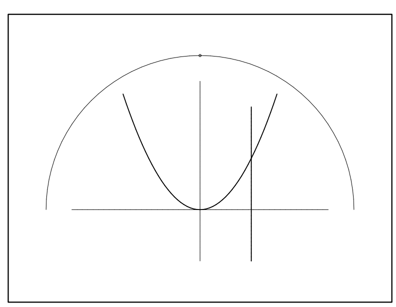


$Y = X + Z$
$Y = X$

Example 2. The homogenized system for the parabola $y = x^2$ and the line $x = 1$ yields the reduced system $Z(Y - Z) = 0$ and $X = Z$, which has two solutions corresponding to the projective points $[1, 1, 1]$ and $[0, 1, 0]$. The point $[1, 1, 1]$ is recognized as the projective version of the affine point $(1, 1)$, and the point $[0, 1, 0]$ is the point at infinity contained in a vertical line.

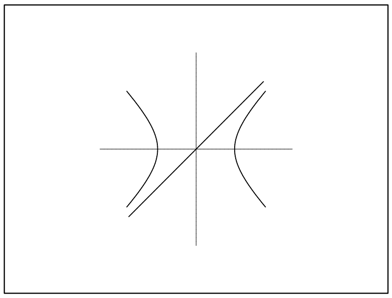


$y = x^2$
$x = 1$

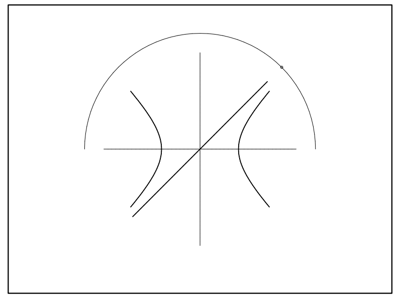


$YZ = X^2$
$X = Z$

Example 3. In this last example, there are no affine points of intersection, and the homogenized system reduces to $Z^2 = 0$ and $X = Y$. This has a single solution $[1, 1, 0]$, and we will verify later that the intersection multiplicity at this point is 2 (which we might guess from the fact that $Z^2 = 0$ has a root of multiplicity 2).



$x^2 - y^2 = 1$
$y = x$



$X^2 - Y^2 = Z^2$
$Y = X$

Bézout's Theorem and Some Examples

With an understanding of the projective plane and the homogenization process, we are ready to give a precise statement of Bézout's Theorem.

Theorem 1 (Bézout's Theorem) *If C and D are complex projective (algebraic) curves with no common components, then*

$$\sum_{P \in C \cap D} i(C \cap D, P) = (\deg C)(\deg D), \tag{1}$$

where $i(C \cap D, P)$ is the intersection multiplicity of C and D at point P.

One of the truly amazing things about our discussion thus far is that by moving to the projective plane and forcing equality instead of inequality for two curves of degree 1 (i.e., forcing two distinct lines to intersect exactly once), we get equality instead of inequality for *any* two curves with no common components.

Of course, our last sentence begs the question of what the assumption of no common components is doing in the theorem statement. So far, we have made no mention of common components. The problem is simple (and easily handled): two copies of the same line intersect at infinitely many points, and we want to eliminate such cases. When the degrees of curves are allowed to be greater than one, it is possible for two curves to have a common component without being identical. For instance, the two curves $C : f(x, y) = xy$ and $D : g(x, y) = x^2 - xy$ are not identical, but they have a common component. The curve C consists of the union of the two lines $x = 0$ and $y = 0$, while the curve D consists of the union of the two lines $x = 0$ and $y = x$. Thus the line $x = 0$ is a component of both curves, and there are infinitely many points in the intersection of C and D. So we see that the assumption about no common components is necessary in the statement of the theorem to avoid the left hand side of Equation (1) being infinite.

There is just one remaining problem with our wonderful theorem: the definition of intersection multiplicity. For a one-variable polynomial $f(x)$, we can essentially use the Fundamental Theorem of Algebra to define the multiplicity of a root. By factoring $f(x)$ into a product of a constant and monic linear factors, we can determine the multiplicity of a root r by observing the power on the factor $x - r$ in the factorization of f. As a quick aside, note that although we can often make an educated guess about the multiplicity of a root from the graph of a polynomial (of either one or two variables), it requires a substantial algebraic result to tell us how to find that multiplicity every time, without fail (even for polynomials of one variable). So our next question is how to define the intersection multiplicity for two arbitrary plane curves (that is, for polynomials of two variables). The answer again involves some substantial algebra.

Recall that if a point (X, Y, Z) has $Z \neq 0$, we can think of (X, Y, Z) as the point (x, y) in the affine plane, where $x = X/Z$ and $y = Y/Z$. Thus, for a point (X, Y, Z) with $Z \neq 0$ (a point not at infinity) in the intersection of two algebraic plane curves C and D defined by $F(X, Y, Z) = 0$ and $G(X, Y, Z) = 0$, respectively, we will define the intersection multiplicity $i(C \cap D, P)$ of C and D at point P by the vector space dimension (over $\mathbf{C}$) of the quotient of the ring of rational functions defined at $P = (x, y)$ by the ideal generated by the polynomials defining the curves $\bar{C}$ and $\bar{D}$ in the *affine* plane. More precisely,

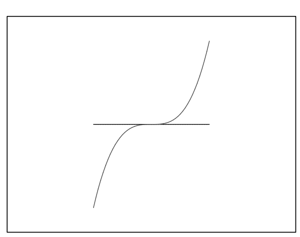
$$i(C \cap D, P) = \dim \frac{\mathcal{O}_P}{(f, g)_P},$$

where the projective curve C defines an affine curve by $F(X/Z, Y/Z, 1) = f(x, y) = 0$, the projective curve D defines an affine curve by $G(X/Z, Y/Z, 1) = g(x, y) = 0$, $\mathcal{O}_P$ is the ring of rational functions defined at P (that is, $\mathcal{O}_P = \{\psi \in \mathbf{C}[x, y] \mid \psi(P) \text{ is defined}\}$), and $(f, g)_P$ is the ideal generated by f and g in $\mathcal{O}_P$.

For points (X, Y, Z) of $C \cap D$ at infinity, we still know that at least one of X and Y is non-zero, so we can dehomogenize with respect to X or Y instead of Z, and define intersection multiplicity in an analogous way to that described above.

While it is not immediately obvious how this definition of intersection multiplicity parallels the concept of multiplicity of a root for a single-variable polynomial, looking at a few examples will illustrate the connection.

Example 4. Let's return to one of our simplest examples: the intersection of $y = x^3$ and $y = 0$.



The only point of intersection is at $(0,0)$, and we do not need to consider points at infinity in this case. Of course, we *know* the correct intersection multiplicity at $(0,0)$ is 3, but we will verify this fact with the intersection multiplicity formula. Some basic ring theory (including localization) is necessary here. First, letting $f(x,y) = y - x^3$ and $g(x,y) = y$, we have

$$\frac{\mathcal{O}_P}{(f,g)_P} \cong \frac{\mathbf{C}[x,y]_{(x,y)}}{(y-x^3,y)_{(x,y)}} \cong \left(\frac{\mathbf{C}[x,y]}{(y-x^3,y)}\right)_{(x,y)} \cong \left(\frac{\mathbf{C}[x]}{(x^3)}\right)_{(x)}.$$

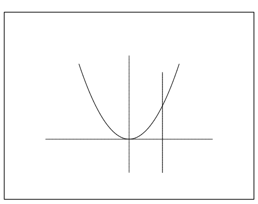
The second isomorphism is a property of localization, and the third isomorphism is the natural one that results from taking the quotients of the ring $\mathbf{C}[x,y]$ and the ideal $(y-x^3,y)$ by the ideal (y). It is easy to verify that $\{1,x,x^2\}$ is a $\mathbf{C}$-basis for $\left(\frac{\mathbf{C}[x]}{(x^3)}\right)_{(x)}$, so $i(C\cap D,P) = 3$.

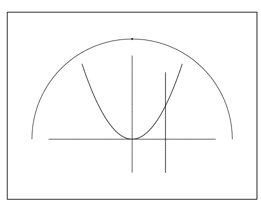
Example 5. Next, let's look at a slightly more interesting example, namely one where there is a point of intersection at infinity. Consider the intersections of $y = x^2$ and $x = 1$. There is only one point of intersection in the affine plane, namely the point $P = (1,1)$. We can compute the intersection multiplicity for this point as in the last example. We have

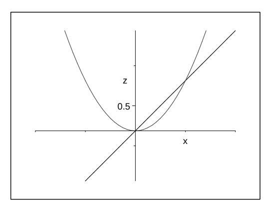
$$\frac{\mathcal{O}_P}{(f,g)_P} \cong \frac{\mathbf{C}[x,y]_{(x-1,y-1)}}{(y-x^2,x-1)_{(x-1,y-1)}} \cong \left(\frac{\mathbf{C}[x,y]}{(y-x^2,x-1)}\right)_{(x-1,y-1)} \cong \left(\frac{\mathbf{C}[y]}{(y-1)}\right)_{(y-1)} \cong \mathbf{C},$$

and since $\mathbf{C}$ is a one-dimensional vector space over itself, $i(C\cap D,P) = 1$.

To find the other point of intersection (the one at infinity), we homogenize the system and consider the projective curves given by $YZ = X^2$ and $X = Z$. Solving this system gives two points, $[1,1,1]$ (the projectivized version of $(1,1)$) and $[0,1,0]$ (the point 'at infinity'). To compute the intersection multiplicity of the latter point, we must dehomogenize with respect to a variable other than Z. Since the variable must be non-zero, we are forced to choose Y, and setting $Y = 1$ gives the dehomogenized system $z = x^2$ and $x = z$, with intersection point $Q = (0,0)$.







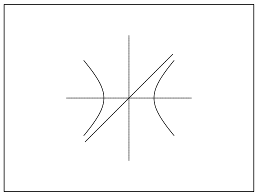
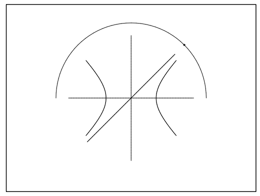
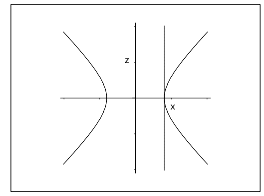
$y = x^2$	$\to$	$YZ = X^2$	$\to$	$z = x^2$
$x = 1$	homogenize with respect to Z	$X = Z$	dehomogenize with $Y = 1$	$x = z$
$P = (1,1)$		$[1,1,1]$ and $[0,1,0]$		$Q = (0,0)$

In this case we have

$$\frac{\mathcal{O}_Q}{(f,g)_Q} \cong \frac{\mathbf{C}[x,z]_{(x,z)}}{(z-x^2,x-z)_{(x,z)}} \cong \left(\frac{\mathbf{C}[x,z]}{(z-x^2,x-z)}\right)_{(x,z)} \cong \left(\frac{\mathbf{C}[x]}{(x(1-x))}\right)_{(x)},$$

and again $i(C \cap D, Q) = 1$.

Example 6. As a final example, we consider the intersection of the hyperbola $x^2 - y^2 = 1$ and one of its asymptotes $y = x$. There are no points of intersection in the natural affine setting, so we must homogenize to get the system $X^2 - Y^2 = Z^2$ and $Y = X$, and then dehomogenize in a different variable. In this case, dehomogenizing by setting either $X = 1$ or $Y = 1$ will work; we proceed by setting $Y = 1$.

				
$x^2 - y^2 = 1$	$\rightarrow$	$X^2 - Y^2 = Z^2$	$\rightarrow$	$x^2 - 1 = z^2$
$y = x$	homogenize with respect to Z	$Y = X$	dehomogenize with $Y = 1$	$1 = x$
		$[1, 1, 0]$		$P = (1, 0)$

Applying our intersection multiplicity formula to the curves defined by $x^2 - 1 = z^2$ and $x = 1$, we have

$$\frac{\mathcal{O}_P}{(f,g)_P} \cong \frac{\mathbf{C}[x,z]_{(x-1,z)}}{(x^2 - 1 - z^2, x - 1)_{(x-1,z)}} \cong \left(\frac{\mathbf{C}[x,z]}{(x^2 - 1 - z^2, x - 1)} \right)_{(x-1,z)} \cong \left(\frac{\mathbf{C}[z]}{(z^2)} \right)_{(z)},$$

and we see that $i(C \cap D, P) = 2$.

You have probably noted a pattern that seems familiar: in order to determine the multiplicity of a point of intersection, first simplify the expression of the local ring of functions. In the factor ideal of the quotient ring, there will be a product of linear factors, and the power of the factor corresponding to the ideal at which the ring is localized will be the intersection multiplicity you are seeking. This is very similar to our process of determining the multiplicity of a root of a polynomial in one variable: factor and observe the power of the factor corresponding to the root you would like to find the multiplicity of.

Final Comments

The analogy between the multiplicity of a root of a single variable polynomial and the intersection multiplicity of a point of two curves is geometrically intuitive. Like many ideas in algebraic geometry, in order to be made precise, both of these concepts require purely algebraic definitions. The geometry precedes algebraic precision, both intuitively and historically (see [D] for a nice historical presentation), but once the algebraic definitions are established, we can see the analogs in the algebra almost as clearly as those in the geometry.

Perhaps the most surprising thing about Bézout's theorem for curves is that its statement is so simple: two projective curves of degrees m and n, with no common components, will share mn points, when counted with multiplicity. After looking at the problem posed by parallel lines, it is remarkable that 'fixing' the non-intersection problem for lines in a minimal way actually 'fixes' the non-intersection problem for any two curves with no common components.

For a proof of Bézout's Theorem, you can consult almost any introductory algebraic geometry text. Although you will find varying degrees of sophistication and generality, because of the algebraic definition of intersection multiplicity, proofs of Bézout's theorem are (not surprisingly) largely algebraic. An appendix in [ST] gives nice development of the projective plane (more detailed than that above) and a relatively elementary outline of a proof of Bézout's theorem.

References

[D] Jean Alexandre Dieudonné, *History of algebraic geometry : an outline of the history and development of algebraic geometry*, Monterey, Calif.: Wadsworth Advanced Books & Software, 1985.

[ST] Joseph H. Silverman and John Tate, *Rational Points on Elliptic Curves*, New York: Springer-Verlag, 1992.