1. Basic Counting Problems

Apollonius was perhaps the greatest of the Greek geometers. He lived around 200 BC. One of his most famous results is:

The locus of points, whose distances from two fixed points are in a constant ratio, is a circle.

This is one of the harder results one can prove using only classical geometry. However using just a little algebra makes the result almost completely trivial.

Change coordinates so that the first point A is at the origin and B lies on the x-axis,

$$A = (0, 0) \qquad\qquad B = (a, 0).$$

Let $P = (x, y)$ be a general point of the locus described above. Let m be the fixed multiple. By assumption

$$|PA| = m|PB|.$$

So we have

$$|PA|^2 = x^2 + y^2 \qquad\qquad |PB|^2 = (x - a)^2 + y^2.$$

Thus

$$x^2 + y^2 = m^2(x - a)^2 + m^2y^2.$$

After a little manipulation we get

$$(x - A)^2 + y^2 = A^2 - B$$

for appropriate A and B, a circle.

There are two general principles to be gleaned from this example:

- To solve problems in analytic geometry, a little bit of algebra goes a long way.
- Since we introduce coordinates to solve this problem, we get to choose where to put the origin.

Definition 1.1. *Let K be a field.* ***Affine n-space over*** *K, denoted $\mathbb{A}^n_K$, is a copy of a vector space V of dimension n. A subset Λ is an* ***affine linear subspace*** *is the translate of a linear subspace of V.*

In other words, affine space is nothing more than a vector space without a preferred point and a line in the affine plane is what a calculus student would call a line but not an undergraduate math major. We will invariably drop the word affine.

Let us now ask our first enumerative geometry question:

Question 1.2. *How many circles pass through three points which are not collinear?*

The answer is one. Probably the easiest way to see this is to use synthetic geometry. Suppose the points are p, q and r. Let L and M be the bisectors of the two lines $\langle p, q\rangle$ and $\langle q, r\rangle$. Then it is easy to see that the point of intersection $L \cap M$ is the centre of the circle we are looking for and that this is the only circle through p, q and r.

Definition 1.3. *Let X be a subset of $\mathbb{A}^n$. The **span** of X, denoted $\langle X\rangle$ is the smallest linear subspace that contains X.*

However there are two entirely different ways to proceed, both of which will prove more fruitful, as they are more general.

Here is the first. Imagine moving the points around. Clearly the answer won't change (or better, if it did the original question does not really make sense). Now suppose that the points become collinear. In this case the only circle through these points is the straight line (a circle of infinite radius) containing them. Supposing that the answer does not change the answer must then be one in general. It is convenient to state more clearly the underlying assumption.

Principle 1.4. *(Principle of continuity) If we are given a problem in enumerative geometry, then the number of solutions is invariant under a continuous change of parameters.*

This is a very useful principle; unfortunately as stated it is clearly false, as there are some obvious counterexamples. The point is to change the definitions, so that this principle holds.

Question 1.5. *In how many points do two lines intersect?*

At first sight the answer would seem to be one; unfortunately some lines are parallel. In fact it is clear that the principle of continuity fails as well.

Example 1.6. *Let L be the line $x = 0$ and let M_t be the line $x = ty+1$, where $t \in K$. Then as t approaches zero, M approaches a line parallel to L, so that the number of points*

$$L \cap M_t,$$

is not constant.

Consider how the principle of continuity fails in this case. We have a sequence of points, $L \cap M_t$, without a limit. If we had a topological space (for example take $K = \mathbb{R}$), then this can only happen if the space is not compact. So we could fix the problem if we can compactify $\mathbb{A}^2$, by adding some points at infinity.

Definition 1.7. *Let K be a field.* ***Projective space of dimension*** *n, denoted $\mathbb{P}^n_K$ or $\mathbb{P}(V)$, is by definition the space of lines in a vector space V of dimension $n+1$.*

Note that, as V is a vector space and not affine space, a line in V contains the origin.

Let us examine this definition more closely. Let V be a vector space of dimension $n+1$. Pick $v \in V - \{0\}$. Then v determines a line $\langle v \rangle$, in the usual way. On the other hand, if w is another non-zero vector, proportional to v, that is $w = \lambda v$, for some $\lambda \neq 0 \in K$, then $\langle v \rangle = \langle w \rangle$. Thus we have proved:

Definition-Lemma 1.8. *Let V be a vector space. $\mathbb{P}(V)$ is equal to the set of points V, modulo the equivalence relation $\sim$, defined as $v \sim w$ iff $v = \lambda w$, $\lambda \in K^*$.*

The equivalence class of the vector v is denoted $[v]$.

Let us see what happens for small values of n. If $n+1 = 0$, then V does not contain any non-zero vectors, and so $\mathbb{P}^{-1}$ is empty. If $n+1 = 1$, then V contains a unique line and so $\mathbb{P}^0$ is a point.

The first interesting case is $\mathbb{P}^1$. Let $V = K^2$. Then $\mathbb{P}^1$ is the set of lines in the plane K^2. Suppose that $v = (X, Y) \in K^2 - \{0\}$. We denote the corresponding point of $\mathbb{P}^1$, by $[v] = [X : Y]$. Then the line spanned by v has a slope, provided $X \neq 0$, and this uniquely determines the line.

The slope $m = Y/X$ takes any value in K. Thus $\mathbb{A}^1 \subset \mathbb{P}^1$. On the other hand, we are only missing one point, corresponding to the line with slope infinity. Thus $\mathbb{P}^1 = \mathbb{A}^1 \cup \{p\}$, and we have compactified $\mathbb{A}^1$, by adding a single point. In fact, we sometimes refer to p as the point at infinity and even denote it by ∞ (the value of Y/X as it were). As an equivalence class, $p = [0 : 1]$.

Note that this situation is completely symmetric. Instead of looking at $y = Y/X$, we could consider $x = X/Y$. In this case we compactify $\mathbb{A}^1$, with coordinate x, by adding the point $q = [1 : 0]$.

It is useful to introduce some more notation to handle this. We denote by U_0 the locus of points of $\mathbb{P}^1$ where $X \neq 0$. As we have already seen, U_0 is a copy of $\mathbb{A}^1$. In this case $\mathbb{P}^1 = U_0 \cup \{[0 : 1]\}$.

Similarly we denote by U_1 the locus of points where $Y \neq 0$. Thus $\mathbb{P}^1 = U_1 \cup \{[1 : 0]\}$. The two sets U_0 and U_1 obviously intersect, along the locus $XY \neq 0$.

Let us see what happens for $\mathbb{P}^2$. Introduce coordinates (X, Y, Z) on $V \simeq K^3$. There are three obvious loci to consider, $X \neq 0$, $Y \neq 0$ and $Z \neq 0$. These induce three subsets of $\mathbb{P}^2$, U_0 and U_1 and U_2. I claim that that U_i is a copy of $\mathbb{A}^2$.

It is easy to see this algebraically. If $(X, Y, Z) \in K^3$ and $X \neq 0$, then $[X : Y : Z] = [1 : Y/X : Z/X]$. Thus the ratios $y = Y/X$ and $z = Z/X$ define coordinates on U_0 and identify U_0 with $\mathbb{A}^2$.

One can also see this geometrically. Any line through the origin of K^3 is determined by its intersection with the locus $X = 1$ (assuming it does intersect, that is assuming the line lies in U_0). But the locus $X = 1$ is surely a copy of $\mathbb{A}^2$.

What is missing? In other words, what is $\mathbb{P}^2 - U_0$? This is the set of points with zero first coordinate, in other words all points of the form $[0 : Y : Z]$. But this is surely a copy of $\mathbb{P}^1$.

In other words we can compactify $\mathbb{A}^2$ by adding a copy of $\mathbb{P}^1$, to get $\mathbb{P}^2$. This copy of $\mathbb{P}^1$ is sometimes called the *line at infinity.*

As before, the situation is completely symmetric. Moreover, all of this generalises in an obvious fashion.

Definition 1.9. *Pick coordinates $X_0, X_1, \dots, X_n$ on K^{n+1}. We will refer to $[X_0 : X_1 : \dots : X_n]$ as* ***homogeneous coordinates*** *on $\mathbb{P}^n$. The subsets $U_0, U_1, \dots, U_n$, given as $X_i \neq 0$, which are copies of $\mathbb{A}^n$, are called the* ***standard open affine subsets****. Indeed the ratios $x_i = \frac{X_j}{X_i}$ define coordinates $x_1, x_2, \dots, \hat{x}_i, \dots, x_n$ on U_i.*

The locus $X_i = 0$ is called the ***hyperplane at infinity****. $\mathbb{P}^n = U_i \cup \{X_i = 0\}$.*

Note that what is at infinity, depends on our point of view. Note also that the term homogeneous coordinates is a bit of a misnomer. In fact $X_0, X_1, \dots, X_n$ are not functions at all, since they are not invariant under rescaling. The only thing that does make sense, is to ask where they are zero (which is invariant under rescaling).

Definition 1.10. *A subset Λ of a projective space $\mathbb{P}(V)$ is called* ***linear*** *if it is given as $\mathbb{P}(W)$, where $W \subset V$ is a linear subspace. The* ***dimension*** *of Λ is the dimension of W minus one.*

In other words a line l in $\mathbb{P}^2$ is the same as a plane W in the corresponding three dimensional vector space K^3.

Lemma 1.11. *Let Λ_1 and Λ_2 be two linear subspaces of $\mathbb{P}^n$ of dimension r and s.*

Then the dimension of the intersection is at least $r + s - n$.

Proof. Let W_1 and W_2 be the corresponding linear subspaces of V, where $\mathbb{P}^n = \mathbb{P}(V)$. Then W_1 has dimension $r + 1$, W_2 has dimension $s + 1$ and V has dimension $n + 1$.

Clearly $\Lambda_1 \cap \Lambda_2 = \mathbb{P}(W_1 \cap W_2)$. On the other hand

$$\begin{aligned}\dim(W_1 \cap W_2) &\geq (r+1) + (s+1) - (n+1) \\ &= r + s - n + 1.\end{aligned}$$

□

The following example shows that we have fixed out problem concerning parallel lines.

Example 1.12. *Let l_1 and l_2 be two lines in $\mathbb{P}^2$. Then $l_1 \cap l_2$ intersect. Indeed the dimension of the intersection is at least zero ($= 1 + 1 - 2$) and the empty set has dimension -1.*

We will see later what happens when we take two parallel lines in $\mathbb{A}^2$ and compactify to $\mathbb{P}^2$. In practice it is often more efficient to work with the codimension and not the dimension.

Definition 1.13. *Let $\Lambda \subset \mathbb{P}^n$ be linear subspace. The* ***codimension*** *of Λ is equal to the difference $n - d$, where d is the dimension of Λ.*

The following is a simple restatement of (1.11); its virtue lies in the fact that is easier to remember and apply:

Lemma 1.14. *Let Λ_1 and Λ_2 be two linear subspaces of $\mathbb{P}^n$ of codimension r and s.*

Then the codimension of the intersection is at most $r+s$. That is the codimension of the intersection is at most the sum of the codimensions.

Let us go back to the principle of continuity. Unfortunately there is another problem.

Question 1.15. *In how many points do a line and a circle meet?*

Example 1.16. *Let L be the line $x = \sqrt{2}$ and C the circle $x^2 + y^2 = 1$ in $\mathbb{A}^2_{\mathbb{R}}$. Then L and C don't intersect.*

Now consider the family of lines L_t, $x = t$. Then $L_t \cap C$ depends on $t \in \mathbb{R}$. If $|t| < 1$ we get two points, if $t = \pm 1$, we get one and if $|t| > 1$ none at all. Thus the principle of continuity does not hold up.

The important thing to realise is that the problem here has nothing to do with the points of intersection moving off to infinity. The problem is that $\mathbb{R}$ is not algebraically closed, so that the equation $y^2 = -1$ has no solutions.

The solution is simple, we should replace $\mathbb{R}$ with $\mathbb{C}$. Now we always get two points (ignoring the possibility that $t = \pm 1$, which we will come back to), $x = t$, $y = \pm\sqrt{1 - t^2}$.

In practice we will work almost exclusively over an algebraically closed field of characteristic zero, which for all intents and purposes means we work over $\mathbb{C}$. In fact working with other fields normally poses extra technical problems, so that working over $\mathbb{C}$ is the most convenient.

We want to talk about curves in $\mathbb{P}^2$. For that we need to look at polynomials. The problem is that polynomials in X, Y and Z don't define functions on $\mathbb{P}^2$, since polynomials are not invariant under rescaling. However, we don't really care what the value of the polynomial is, all we care about is whether or not the polynomial is zero.

Definition 1.17. *Let $F(X) \in K[X]$ be a polynomial in the variables $X_0, X_1, \dots, X_n$. We say that F is **homogeneous** if every non-zero term of F has the same degree d.*

Lemma 1.18. *Let $F(X) \in K[X]$.*

1. *If F is homogeneous of degree d, then $F(\lambda X) = \lambda^d F(X)$, for all $\lambda \in K$ (we adopt the convention here that $0^0 = 1$).*
2. *Conversely, if $F(\lambda X) = \lambda^d F(X)$, for all $\lambda \in K$ and K is infinite then F is homogeneous of degree d.*

Proof. (1) is clear. Suppose now that F is any polynomial. Then $F = \sum_i F_i$ has a unique decomposition, where F_i is homogeneous of degree i. If $F(\lambda X) = \lambda^d F(X)$, then this forces $\lambda^i F_i(X) = \lambda^d F_i(X)$, for every i. If K is infinite then for every $i \neq d$, we can pick λ, so that $\lambda^i \neq \lambda^d$. Thus $F_i(X) = 0$, for all $i \neq d$. $\square$

Definition 1.19. *Let $X \subset \mathbb{A}^n$ be a subset. We say that X is a **closed subvariety** of $\mathbb{A}^n$ if the set X is equal to the zero locus of a collection of polynomials.*

*Let $X \subset \mathbb{P}^n$ be a subset. We say that X is a **closed subvariety** of $\mathbb{P}^n$ if the set X is equal to the zero locus of a collection of homogenous polynomials.*

Example 1.20. *The circle $x^2 + y^2 = 1$ is an affine subvariety. It is the zero locus of the polynomial $x^2 + y^2 - 1$.*

Lemma 1.21. *$\Lambda \subset \mathbb{P}^n$ is a linear subspace iff it is defined by a collection of homogeneous linear equations.*

In particular every linear subspace of $\mathbb{P}^n$ is a projective subvariety.

Proof. Clear, since a subset $W \subset V$ is a linear subspace iff it is defined by homogeneous linear equations. $\square$

One of the key points, is that we can go backwards and forwards between affine and projective varieties.

First let us suppose that we are given a subset $V \subset \mathbb{P}^n$. Clearly we can form $V \subset U_0$ simply by intersecting V with U_0. Suppose that V is a closed subvariety, say defined by $F_\alpha(X)$ homogenous. Define $f_\alpha(x)$ by replacing X_i by X_i/X_0. it is pretty easy to see that V_0 is defined by the f_α.

Conversely suppose we are given f_α, which defines $V \subset \mathbb{A}^n$. Then we can form $F_\alpha(X)$ homogeneous, simply by topping up each term of f_α, by the appropriate power of X_0. This defines $\bar{V}$ in X. Again, it is not hard to see that $\bar{V} \cap U_0 = V$.

In both cases, the best way to see what is going on, is to look at some examples.

Example 1.22. *Suppose we consider $x^2 + y^2 = 1$ inside $\mathbb{A}^2$. We may think of this as $U_2 \subset \mathbb{P}^3$, with coordinates X, Y and Z. Replace $x^2+y^2 = 1$ by x^2+y^2-1. This has degree two. The first two terms have degree two, and there is nothing to do (apart from replacing lower caps by upper). The last term has degree zero. To make this homogeneous then, we need to multiply by Z^2. We get $X^2 + Y^2 - Z^2$. Now suppose we want to work on U_0. Then we divide through by X^2 and replace Y/X by y and Z/X by z, to get $1 + y^2 - z^2$. Note the quick way to do this is simply to replace X by 1 and replace upper caps by lower.*

Consider $y = x^3$. We get $y - x^3$. Consider this inside $\mathbb{P}^2$, with coordinates X, Y and Z. We get $YZ^2 - X^3$. Now work inside U_1. We get $z^2 - x^3$.

It is interesting to see what happens to parallel lines in $\mathbb{A}^2$. Let L be the line $x = 0$ and let M_t be the line $x = ty + 1$, where $t \in K$. Then L becomes the line $X = 0$ and M_t the line $X = tY + Z$. When $t = 0$, we get $X = Z$. Thus $Z = 0$, and we get the point $[0 : 1 : 0]$. Thus our two parallel lines intersect along the line at infinity, at the point $[0 : 1 : 0]$, corresponding to the fact that both lines are horizontal.

In fact it is interesting to consider the family in the coordinate patch $Y \neq 0$. We get $x = 0$ and $x = t + z$, which is equivalent to $x = 0$ and $z = -t$.

Note that these processes are not quite inverse. Suppose we start with $X = 0$ inside $\mathbb{P}^2$. If we go to the coordinate patch U_0 then we get the empty set. Going back to $\mathbb{P}^2$, we get the empty set. The whole point is that the whole of $X = 0$ completely avoids the set U_0.