5. Smoothness and the Zariski tangent space

We want to give an algebraic notion of the tangent space. In differential geometry, tangent vectors are equivalence classes of maps of intervals in $\mathbb{R}$ into the manifold. This definition lifts to algebraic geometry over $\mathbb{C}$ but not over any other field (for example a field of characteristic p).

Classically tangent vectors are determined by taking derivatives, and the tangent space to a variety X at x is then the space of tangent directions, in the whole space, which are tangent to X. Even is this is how we will compute the tangent space in general, it is still desirable to have an intrinsic definition, that is, a definition which does not use the fact that X is embedded in $\mathbb{P}^n$.

Now note first that the notion of smoothness is surely local and that if we want an intrinsic definition, then we want a definition that only uses the functions on X. Putting this together, smoothness should be a property of the local ring of X at p. On the other hand taking derivatives is the same as linear approximation, which means dropping quadratic and higher terms.

Definition 5.1. *Let X be a variety and let $p \in X$ be a point of X. The **Zariski tangent space** of X at p, denoted T_pX, is equal to the dual of the quotient*

$$\mathfrak{m}/\mathfrak{m}^2,$$

where $\mathfrak{m}$ is the maximal ideal of $\mathcal{O}_{X,p}$.

Note that $\mathfrak{m}/\mathfrak{m}^2$ is a vector space. Suppose that we are given a morphism

$$f\colon X \longrightarrow Y,$$

which sends p to q. In this case there is a ring homomorphism

$$f^*\colon \mathcal{O}_{Y,q} \longrightarrow \mathcal{O}_{X,p}$$

which sends the maximal ideal $\mathfrak{n}$ into the maximal ideal $\mathfrak{m}$. Thus we get an induced map

$$df\colon \mathfrak{n}/\mathfrak{n}^2 \longrightarrow \mathfrak{m}/\mathfrak{m}^2$$

On the other hand, geometrically the map on tangent spaces obviously goes the other way. Therefore it follows that we really do want the dual of $\mathfrak{m}/\mathfrak{m}^2$. In fact $\mathfrak{m}/\mathfrak{m}^2$ is the dual of the Zariski tangent space, and is referred to as the *cotangent space*.

In particular, given a morphism $f\colon X \longrightarrow Y$ carrying p to q, then there is a linear map

$$df\colon T_pX \longrightarrow T_qY.$$

Definition 5.2. *Let X be a quasi-projective variety.*

*We say that X is **smooth** at p if the local dimension of X at p is equal to the dimension of the Zariski tangent space at p.*

Now the tangent space to $\mathbb{A}^n$ is canonically a copy of $\mathbb{A}^n$ itself, considered as a vector space based at the point in question. If $X \subset \mathbb{A}^n$, then the tangent space to X is included inside the tangent space to $\mathbb{A}^n$. The question is then how to describe this subspace.

Lemma 5.3. *Let $X \subset \mathbb{A}^n$ be an affine variety. Suppose that $f_1, f_2, \dots, f_k$ generate the ideal I of X. Then the tangent space of X at p, considered as a subspace of the tangent space to $\mathbb{A}^n$, via the inclusion of X in $\mathbb{A}^n$, is equal to the kernel of the Jacobian matrix.*

Proof. Clearly it is easier to give the dual description of the cotangent space.

If $\mathfrak{m}$ is the maximal ideal of $\mathcal{O}_{\mathbb{A}^n,p}$ and $\mathfrak{n}$ is the maximal ideal of $\mathcal{O}_{X,p}$, then clearly the natural map $\mathfrak{m} \longrightarrow \mathfrak{n}$ is surjective, so that the induced map on contangent spaces is surjective. Dually, the induced map on the Zariski tangent space is injective, so that T_pX is indeed included in $T_p\mathbb{A}^n$.

We may as well choose coordinates $x_1, x_2, \dots, x_n$ so that p is the origin. In this case $\mathfrak{m} = \langle x_1, x_2, \dots, x_n \rangle$ and $\mathfrak{n} = \mathfrak{m}/I$. Moreover $\mathfrak{m}/\mathfrak{m}^2$ is the vector space spanned by $dx_1, dx_2, \dots, dx_n$, where dx_i denotes the equivalence class $x_i + \mathfrak{m}^2$, and $\mathfrak{n}/\mathfrak{n}^2$ is canonically isomorphic to $\mathfrak{m}/(\mathfrak{m}^2 + I)$. Now the transpose of the Jacobian matrix, defines a linear map

$$K^k \longrightarrow K^n = T^*_p\mathbb{A}^n,$$

and it suffices to prove that the image of this map is the kernel of the map

$$df \colon \mathfrak{m}/\mathfrak{m}^2 \longrightarrow \mathfrak{n}/\mathfrak{n}^2.$$

Let $g \in \mathfrak{m}$. Then

$$g(x) = \sum a_i x_i + h(x),$$

where $h(x) \in \mathfrak{m}^2$. Thus the image of $g(x)$ in $\mathfrak{m}/\mathfrak{m}^2$ is equal to $\sum_i a_i dx_i$. Moreover, by standard calculus a_i is nothing more than

$$a_i = \left.\frac{\partial g}{\partial x_i}\right|_p .$$

Thus the kernel of the map df is generated by the image of f_i in $\mathfrak{m}/\mathfrak{m}^2$, which is

$$\sum_j \left.\frac{\partial f_i}{\partial x_j}\right|_p dx_j,$$

which is nothing more than the image of the Jacobian. □

Lemma 5.4. *Let X be a quasi-projective variety. Then the function*

$$\lambda\colon X \longrightarrow \mathbb{N},$$

is upper semi-continuous, where $\lambda(x)$ is the dimension of the Zariski tangent space at x.

Proof. Clearly this result is local on X so that we may assume that X is affine. In this case the Jacobian matrix defines a morphism π from X to the space of matrices and the locus where the Zariski tangent space has a fixed dimension is equal to the locus where this morphism lands in the space of matrices of fixed rank. Put differently the function λ is the composition of π and an affine linear function of the rank on the space of matrices. Since the rank function is upper semicontinuous, the result follows. □

Lemma 5.5. *Every irreducible quasi-projective variety is birational to a hypersurface.*

Proof. Let X be a quasi-projective variety of dimension k, with function field L/K. Let $L/M/K$ be an intermediary field, such that M/K is purely transcendental of transcendence degree, so that L/M is algebraic. As L/M is a finitely generated extension, it follows that L/M is finite. Suppose that L/M is not separable. Then there is an element $y \in L$ such that $y \notin M$ but $x_1 = y^p \in M$. We may extend x_1 to a transcendence basis $x_1, x_2, \dots, x_k$ of M/K. Let M' be the intermediary field generated by $y, x_2, x_3, \dots, x_k$. Then M'/K is a purely transcendental extension of K and

$$[L : M] = [L : M'][M' : M] = p[L : M'].$$

Repeatedely replacing M by M' we may assume that L/M is a separable extension.

By the primitive element Theorem, L/M is generated by one element, say α. It follows that there is polynomial $f(x) \in M[x]$ such that α is a root of $f(x)$. If $M = K(x_1, x_2, \dots, x_k)$, then clearing denominators, we may assume that $f(x) \in K[x_1, x_2, \dots, x_k][x] \simeq K[x_1, x_2, \dots, x_{k+1}]$. But then X is birational to the hypersurface defined by $F(X)$, where $F(X)$ is the homogenisation of $f(x)$. □

Proposition 5.6. *The set of smooth points of any variety is Zariski dense.*

Proof. Since the dimension of the Zariski tangent space is upper semicontinuous, and always at least the dimension of the variety, it suffices to prove that every irreducible variety contains at least one smooth point. By (5.5) we may assume that X is a hypersurface. Passing to an affine open subset, we may assume that X is an affine hypersurface. Let f be a definining equation, so that f is an irreducible polynomial. Then the set of singular points of X is equal to the locus of points where every partial derivative vanishes. If g is a non-zero partial derivative of f, then g is a non-zero polynomial of degree one less than f, and so cannot vanish on X.

If all the partial derivatives of f are the zero polynomial, then f is a pth power, where the characteristic is p, which contradicts the fact that f is irreducible. □

A basic result in the theory of C^∞-maps is Sard's Theorem, which states that the set of points where a map is singular is a subset of measure zero (of the base). Since any holomorphic map between complex manifolds is automatically C^∞, and the derivative of a polynomial is the same as the derivative as a holomorphic function, it follows that any morphism between varieties, over $\mathbb{C}$, is smooth over an open subset. In fact by the Lefschetz principle, this result extends to any variety over $\mathbb{C}$.

Theorem 5.7. *Let $f\colon X \longrightarrow Y$ be a morphism of varieties over a field of characteristic zero.*

Then there is a dense open subset U of Y such that if $q \in U$ and $p \in f^{-1}(q) \cap X_{sm}$ then the differential $df_p\colon T_pX \longrightarrow T_qY$ is surjective. Further, if X is smooth, then the fibres $f^{-1}(q)$ are smooth if $q \in U$.

Let us recall the Lefschetz principle. First recall the notion of a first order theory of logic. Basically this means that one describes a theory of mathematics using a theory based on predicate calculus. For example, the following is a true statement from the first order theory of number theory,

$$\forall n \forall x \forall y \forall z \quad n \geq 3 \implies x^n + y^n \neq z^n.$$

One basic and desirable property of a first order theory of logic is that it is complete. In other words every possible statement (meaning anything that is well-formed) can be either proved or disproved. It is a very well-known result that the first order theory of number theory is not complete (Gödel's Incompleteness Theorem). What is perhaps more surprising is that there are interesting theories which are complete.

Theorem 5.8. *The first order logic of algebraically closed fields of characteristic zero is complete.*

Notice that a typical statement of the first order logic of fields is that a system of polynomial equations does or does not have solution. Since most statements in algebraic geometry turn on whether or not a system of polynomial equations have a solution, the following result is very useful.

Principle 5.9 (Lefschetz Principle)**.** *Every statement in the first order logic of algebraically closed fields of characteristic zero, which is true over $\mathbb{C}$, is in fact true over any algebraically closed field of characteristic.*

In fact this principle is immediate from (5.8). Suppose that p is a statement in the first order logic of algebraically closed fields of characteristic zero. By completeness, we can either prove p or not p. Since p holds over the complex numbers, there is no way we can prove not p. Therefore there must be a proof of p. But this proof is valid over any field of characteristic zero, so p holds over any algebraically closed field of characteristic zero.

A typical application of the Lefschetz principle is (5.7). By Sard's Theorem, we know that (5.7) holds over $\mathbb{C}$. On the other hand, (5.7), can be reformulated in the first order logic of algebraically closed fields of characteristic zero. Therefore by the Lefschetz principle, (5.7) is true over algebraically closed field of characteristic zero.

Perhaps even more interesting, is that (5.7) fails in characteristic p. Let $f\colon \mathbb{A}^1 \longrightarrow \mathbb{A}^1$ be the morphism $t \longrightarrow t^p$. If we fix s, then the equation

$$x^p - s$$

is purely inseparable, that is, has only one root. Thus f is a bijection. However, df is the zero map, since $dz^p = pz^{p-1}dz = 0$. Thus df_p is nowhere surjective. Note that the fibres of this map, as schemes, are isomorphic to zero dimensional schemes of length p.

We now want to aim for a version of the Inverse function Theorem. In differential geometry, the inverse function theorem states that if a function is an isomorphism on tangent spaces, then it is locally an isomorphism. Unfortunately this is too much to expect in algebraic geometry, since the Zariski topology is too weak for this to be true. For example consider a curve which double covers another curve. At any point where there are two points in the fibre, the map on tangent spaces is an isomorphism. But there is no Zariski neighbourhood of any point where the map is an isomorphism.

Thus a minimal requirement is that the morphism is a bijection. Note that this is not enough in general for a morphism between algebraic varieties to be an isomorphism. For example in characteristic

p, Frobenius is nowhere smooth and even in characteristic zero, the parametrisation of the cuspidal cubic is a bijection but not an isomorphism.

Lemma 5.10. *If $f\colon X \longrightarrow Y$ is a projective morphism with finite fibres, then f is finite.*

Proof. Since the result is local on the base, we may assume that Y is affine. By assumption $X \subset Y \times \mathbb{P}^n$ and we are projecting onto the first factor. Possibly passing to a smaller open subset of Y, we may assume that there is a point $p \in \mathbb{P}^n$ such that X does not intersect $Y \times \{p\}$.

As the blow up of $\mathbb{P}^n$ at p, fibres over $\mathbb{P}^{n-1}$ with fibres isomorphic to $\mathbb{P}^1$, and the composition of finite morphisms is finite, we may assume that $n = 1$, by induction on n.

We may assume that p is the point at infinity, so that $X \subset Y \times \mathbb{A}^1$, and X is affine. Now X is defined by $f(x) \in A(Y)[x]$, where the coefficients of $f(x)$ lie in $A(Y)$. Suppose that

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0.$$

We may always assume that a_n does not vanish at y. Passing to the locus where a_n does not vanish, we may assume that a_n is a unit, so that dividing by a_n, we may assume that $a_n = 1$. In this case the ring B is a quotient of the ring

$$A[x]/\langle f\rangle.$$

But the latter is generated over A by 1, x, $\dots x^{n-1}$, and so is a finitely generated module over A. ☐

Theorem 5.11. *Let $f\colon X \longrightarrow Y$ be a projective morphism between quasi-projective varieties.*

Then f is an isomorphism iff it is a bijection and the differential df_p is injective.

Proof. One direction is clear. Otherwise assume that f is projective and a bijection on closed points. Then f is finite by (5.10). The result is local on the base, so we may assume that $Y = \operatorname{Spec} C$ is affine, in which case $X = \operatorname{Spec} D$ is affine, where C is a finitely generated D-module. Pick $x \in X$ and let $y = f(x)$. Then $x = \mathfrak{p}$ and $y = \mathfrak{q}$ are two prime ideals in C and D. Let A be the local ring of Y at y, B of X at x. Then A is the localisation of C at the multiplicative subset $S = C - \mathfrak{q}$ and as x is the unique point of the fibre, B is the localisation of D by the multiplicative subset $T = S \cdot D$, so that B is a finitely generated A-module.

Let $\phi\colon A \longrightarrow B$ be the induced ring homomorphism. Then B is a finitely generated A-module and we just need to show that ϕ is an isomorphism.

As f is a bijection on closed points, it follows that ϕ is injective. So we might as well suppose that ϕ is an inclusion. Let $\mathfrak{m}$ be the maximal ideal of A and let $\mathfrak{n}$ be the maximal ideal of B. By assumption

$$\frac{\mathfrak{m}}{\mathfrak{m}^2} \longrightarrow \frac{\mathfrak{n}}{\mathfrak{n}^2},$$

is surjective. But then

$$\mathfrak{m}B + \mathfrak{n}^2 = \mathfrak{n}.$$

By Nakayama's Lemma applied to the B-module $\mathfrak{n}/\mathfrak{m}B$, it follows that $\mathfrak{m}B = \mathfrak{n}$. But then

$$B/A \otimes A/\mathfrak{m} = B/(\mathfrak{m}B + A) = B/(\mathfrak{n} + A) = 0.$$

Nakayama's Lemma applied to the finitely generated A-module B/A implies that $B/A = 0$ so that ϕ is an isomorphism. $\square$

Lemma 5.12. *Suppose that $X \subset \mathbb{P}^n$ is a quasi-projective variety and suppose that $\pi\colon X \longrightarrow Y$ is the morphism induced by projection from a linear subspace.*

Let $y \in Y$. Then $\pi^{-1}(y) = \langle \Lambda, y\rangle \cap X$. If further this fibre consists of one point, then the map between Zariski tangent spaces is an isomorphism if the intersection of $\langle \Lambda, x\rangle$ with the Zariksi tangent space to X at x has dimension zero.

Proof. Easy. $\square$

Lemma 5.13. *Let X be a smooth irreducible subset of $\mathbb{P}^n$ of dimension k. Consider the projection Y of X down to a smaller dimensional projective space $\mathbb{P}^m$, from a linear space Λ of dimension $n - m - 1$.*

If the dimension of $m \geq 2k + 1$ and Λ is general (that is belongs to an appropriate open subset of the Grassmannian) then π is an isomorphism.

Proof. Since projection from a general linear space is the same as a sequence of projections from general points, we may assume that Λ is in fact a point p, so that $m = n - 1$.

Now we know that π is a bijection provided that p does not lie on any secant line. Since the secant variety has dimension at most $2k + 1$, it follows that we may certainly find a point away from the secant variety, provided that $n > 2k + 1$. Now since a tangent line is a limit of secant lines, it follows that such a point will also not lie on any tangent lines.

But then π is then an isomorphism on tangent spaces, whence an isomorphism. $\square$

For example, it follows that any curve may be embedded in $\mathbb{P}^3$ and any surface in $\mathbb{P}^5$. Now let us turn to the following classical problem in enumerative geometry.

Question 5.14. *Let C be a curve in $\mathbb{P}^2$ and let $p \in \mathbb{P}^2$.*

How many tangent lines does p lie on?

The first thing that we will need is a natty way to describe the projective tangent space to a variety.

Definition 5.15. *Let $X \subset \mathbb{P}^n$.*

*The **projective tangent space to X at** p is the closure of the affine tangent space.*

In other words the projective tangent space has the same dimension as the affine tangent space and is obtained by adding the suitable points at infinity. Suppose that the curve is defined by the polynomial $F(X, Y, Z)$. Then the tangent line to C at p, is

$$\left.\frac{\partial F}{\partial X}\right|_p X + \left.\frac{\partial F}{\partial Y}\right|_p Y + \left.\frac{\partial F}{\partial Z}\right|_p Z.$$

Of course it suffices to check that we get the right answer on an affine piece.

Lemma 5.16. *Let F be a homogeneous polynomial of degree d in $X_0, X_1, \dots, X_n$. Then*

$$dF = \sum X_i \frac{\partial F}{\partial X_i}$$

Proof. Both sides are linear in F. Thus it suffices to prove this for a monomial of degree d, when the result is clear. □

It follows then that the tangent line above does indeed pass through p. The rest is easy.

Finally we will need Bézout's Theorem.

Theorem 5.17 (Bézout's Theorem)**.** *Let C and D be two curves defined by homogenous polynomials of degrees d and e. Suppose that $C \cap D$ does not contain a curve.*

Then $|C \cap D|$ is at most de, with equality iff the intersection of the two tangent spaces at $p \in C \cap D$ is equal to p.

We are now ready to answer (5.14).

Lemma 5.18. *Let $C \subset \mathbb{P}^n$ be a curve in $\mathbb{P}^2$ and let $p \in \mathbb{P}^2$ be a general point.*

Then p lies on $d(d-1)$ tangent lines.

Proof. Fix $p = [a : b : c]$ and let D be the curve defined by

$$G = a\frac{\partial F}{\partial X} + b\frac{\partial F}{\partial Y} + c\frac{\partial F}{\partial Z}.$$

Then G is a polynomial of degree $d-1$. Consider a point q where C intersects D. Then the tangent line to C at q is given by

$$\left.\frac{\partial F}{\partial X}\right|_q X + \left.\frac{\partial F}{\partial Y}\right|_q Y + \left.\frac{\partial F}{\partial Z}\right|_q Z.$$

But then since p satisfies this equation, as q lies on D, it follows that p lies on the tangent line of C at q. Similarly it is easy to check the converse, that if p lies on the tangent line to C at q, then q is an intersection point of C and D.

Now apply Bézout's Theorem. □

There is an interesting way to look at all of this. In fact one may generalise the result above to the case of curves with nodes. Note that if you take a curve in $\mathbb{P}^3$ and take a general projection down to $\mathbb{P}^2$, then you get a nodal curve. Indeed it is easy to pick the point of projection not on a tangent line, since the space of tangent lines obviously sweeps out a surface; it is a little more involved to show that the space of three secant lines is a proper subvariety. (5.18) was then generalised to this case and it was shown that if δ is the number of nodes, then the number

$$\frac{d(d-1)}{2} - \delta$$

is an invariant of the curve.

Here is another way to look at this. Suppose that we project our curve down to $\mathbb{P}^1$ from a point. Then we get a finite cover of $\mathbb{P}^1$, with d points in the general fibre. Lines tangent to C passing through p then count the number of branch points, that is, the number of points in the base where the fibre has fewer than d points. Since this tangent line is only tangent to p and is simply tangent (that is, there are no flex points) there are $d-1$ points in this fibre, and the ramification point corresponding to the branch point is where two sheets come together.

The modern approach to this invariant is quite different. If we are over the complex numbers $\mathbb{C}$, changing perspective, we may view the curve C as a Riemann surface covering another Riemann surface D. Now the basic topological invariant of a compact oriented Riemann surface is it's genus. In these terms there is a simple formula that connects the genus of C and B, in terms of the ramification data, known as Riemann-Hurwitz,

$$2g - 2 = d(2h-2) + b,$$

where g is the genus of C, h the genus of B, d the order of the cover and b the contribution from the ramification points. Indeed if locally on C, the map is given as $z \longrightarrow z^e$ so that e sheets come together, the contribution is $e-1$.

In our case, $B = \mathbb{P}^1$ which is of genus 0, for each branch point, we have simple ramification, so that $e = 2$ and the contribution is one, making a total $b = d(d-1)$. Thus

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

Solving for g we get

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

Note that if $d \leq 2$, then we get $g = 0$ as expected (that is $C \simeq \mathbb{P}^1$) and if $d = 3$ then we get an elliptic curve.

It also seems worth pointing out that if we take a smooth variety X and blow up a point p, then the exceptional divisor E is canonically the projectivisation of the Zariski tangent space to X at p,

$$E = \mathbb{P}(T_p X).$$

Indeed the point is that E picks up the different tangent directions to X at p, and this is exactly the set of lines in $T_p X$. Note the difference between the projective tangent space and the projectivisation of the tangent space.

It also seems worth pointing out that one defines the Zariski tangent space to a scheme X, at a point x, using exactly the same definition, the dual of

$$\mathfrak{m}/\mathfrak{m}^2,$$

where $\mathfrak{m} \subset \mathcal{O}_{X,x}$ is the maximal ideal of the local ring. However in general, if we have the equality of dimensions of both the Zariski tangent space and the local dimension, we only call X **regular at** $x \in X$. Smoothness is a more restricted notion in general.

Having said this, if X is a quasi-projective variety over an algebraically closed field then X is smooth as a variety if and only if it is smooth as a scheme over $\operatorname{Spec} k$. In fact an abstract variety over $\operatorname{Spec} k$ is smooth if and only if it is regular. Note that if x is a specialisation of ξ and X is regular at x then X is regular at ξ, so it is enough to check that X is regular at the closed points.

Note that one can sometimes use the Zariski tangent space to identify embedded points. If X is a scheme and $Y = X_{\text{red}}$ is the reduced subscheme then $x \in X$ is an embedded point if

$$\dim T_x X > \dim T_x Y,$$

and X is reduced away from x. For example, if X is not regular at x but Y is regular at x then $x \in X$ is an embedded point. Note however that it is possible that $x \in X$ is an embedded point but the Zariski tangent space is no bigger than it should be; for example if $\mathcal{H}_0^3$ is the punctual Hilbert scheme of a smooth surface then the underlying variety is a quadric cone. The vertex of the cone corresponds to the unique zero dimensional length three scheme which is not curivilinear and this is an embedded point, even though the Zariski tangent space is three dimensional.

It is interesting to see which toric varieties are smooth. The question is local, so we might as well assume that $X = U_\sigma$ is affine. If $\sigma \subset N_{\mathbb{R}}$ does not span $N_{\mathbb{R}}$, then $X \simeq U_{\sigma'} \times \mathbb{G}_m^l$, where σ' is the same cone as σ embedded in the space it spans. So we might as well assume that σ spans $N_{\mathbb{R}}$. In this case X contains a unique fixed point x_σ which is in the closure of every orbit. Since X only contains finitely many orbits, it follows that X is smooth if and only if X is regular at x_σ. The maximal ideal of x_σ is generated by χ^u, where $u \in S_\sigma$. The square of the maximal is generated by χ^{u+v}, where u and v are two elements of S_σ. So a basis for $\mathfrak{m}/\mathfrak{m}^2$ is given by elements of S_σ that are not sums of two elements. Since the elements of S_σ generate the group M, the elements of S_σ which are not sums of two elements, must generate the group. Given an extremal ray of $\check{\sigma}$, a primitive generator of this ray is not the sum of two elements in S_σ. So $\check{\sigma}$ must have n edges and they must generate M. So these elements are a basis of the lattice and in fact $X \simeq \mathbb{A}_k^n$.