Navigating the motivic world

Daniel Dugger

Department of Mathematics, University of Oregon, Eugene, OR 97403

E-mail address: `ddugger@math.uoregon.edu`

Contents

Introduction

Part 1

Motivation

CHAPTER 1

Introduction to the Weil conjectures

The story of the Weil conjectures has many layers to it. On the surface it might seem simple enough: the conjectures postulated the existence of certain formulas for the number of solutions to equations over finite fields. The rather amazing thing, however, is that the conjectures also provided a link between these formulas and the world of algebraic topology. Understanding why such a link should exist—and in the process, proving the conjectures—was one of the greatest mathematical achievements of the twentieth century. It is also one which has had lasting implications. Work on the Weil conjectures was one of the first places that deep algebraic-topological ideas were developed for varieties over arbitrary fields. The continuation of that development has taken us through Quillen's algebraic K-theory and Voevodsky's motivic cohomology, and is still a very active area of research.

In Sections 1 and 2 of this chapter we will introduce the Weil conjectures via several examples. Section 3 discusses ways in which the conjectures are analogs of properties of the classical Riemann zeta function. Then in Sections 4 and 5 we outline the cohomological approach to the problem, first suggested by Weil and later carried out by Grothendieck and his collaborators. The chapter has two appendices, both dealing with further examples. Appendix A introduces the reader to some tools for computer calculations. Appendix B treats the class of examples originally handled by Weil, which involve an intriguing connection with algebraic number theory.

1. A first look

Let's dive right in. Suppose given polynomials $f_1, \ldots, f_k \in \mathbb{Z}[x_1, \ldots, x_n]$. Fix a prime p, and look at solutions to the equations

$$f_1(x_1, \ldots, x_n) = f_2(x_1, \ldots, x_n) = \ldots = f_k(x_1, \ldots, x_n) = 0$$

where $x_1, \ldots, x_n \in \mathbb{F}_{p^m}$ and the coefficients of the f_i's have been reduced modulo p. Let N_m denote the number of such solutions. Our task will be to develop a formula for N_m as a function of m.

In the language of algebraic geometry, the mod p reductions of the f_i's define an algebraic variety $X = V(f_1, \ldots, f_k)$ over the field $\mathbb{F}_p$. The set of points of this variety defined over the extension field $\mathbb{F}_{p^m}$ is usually denoted $X(\mathbb{F}_{p^m})$, and we have $N_m = \#X(\mathbb{F}_{p^m})$.

Example 1.1. Consider the single equation $y^2 = x^3 + x$ and take $p = 2$. Over $\mathbb{F}_2$ there are exactly two solutions for (x, y), namely $(0, 0)$ and $(1, 0)$. Over $\mathbb{F}_4 = \mathbb{F}_2[\omega]/(\omega^2+\omega+1)$ one has four solutions: $(0, 0)$, $(1, 0)$, (ω, ω), and $(\omega+1, \omega+1)$. So we have $N_1 = 2$ and $N_2 = 4$.

We will mostly want to talk about *projective* varieties rather than affine varieties. If F is a field, let $\mathbb{A}^n(F) = \{(x_1, \ldots, x_n) : x_i \in F\}$ and let $\mathbb{P}^n(F) = [\mathbb{A}^{n+1}(F) - 0]/F^*$ where F^* acts on $\mathbb{A}^n(F)$ by scalar multiplication. Given homogeneous polynomials $f_i \in F[x_0, \ldots, x_n]$, consider the set of common solutions to the f_i's inside of $\mathbb{P}^n(F)$. These are the F-valued points of the projective algebraic variety $X = V(f_1, \ldots, f_k)$.

EXAMPLE 1.2. Consider the single equation $y^2z = x^3 + xz^2$. Over $\mathbb{F}_2$ there are three solutions in $\mathbb{P}^3$, namely $[0, 0, 1]$, $[1, 0, 1]$, and $[0, 1, 0]$. Over $\mathbb{F}_4$ there are five solutions: $[0, 0, 1]$, $[1, 0, 1]$, $[\omega, \omega, 1]$, $[\omega + 1, \omega + 1, 1]$, $[0, 1, 0]$.

Given homogeneous polynomials $f_1, \ldots, f_k \in \mathbb{Z}[x_1, \ldots, x_{n+1}]$, we will be concerned with counting the number of points of $V(f_1, \ldots, f_k)$ defined over $\mathbb{F}_{p^m}$. We will start by looking at two elementary examples which can be completely understood by hand.

EXAMPLE 1.3. $X = \mathbb{P}^d$. As $\mathbb{P}^d(\mathbb{F}_{p^m}) = [\mathbb{A}^{d+1} - 0]/(\mathbb{F}_{p^m})^*$ we have

$$N_m(X) = \frac{(p^m)^{d+1} - 1}{p^m - 1} = 1 + p^m + p^{2m} + \cdots + p^{dm}.$$

EXAMPLE 1.4. $X = \mathrm{Gr}_2(\mathbb{A}^r)$, the variety of 2-planes in $\mathbb{A}^r$. The points of X are linearly independent pairs of vectors modulo the equivalence relation given by the action of GL_2. To specify a linearly independent pair, one chooses a nonzero vector v_1 and then any vector v_2 which is not in the span of v_1. The number of ways to make these choices is $[(p^m)^r - 1] \cdot [(p^m)^r - p^m]$. Similarly, the number of elements of $GL_2(\mathbb{F}_{p^m})$ is $[(p^m)^2 - 1] \cdot [(p^m)^2 - p^m]$. Hence one obtains

$$N_m(X) = \frac{[(p^m)^r - 1] \cdot [(p^m)^r - p^m]}{[(p^m)^2 - 1] \cdot [(p^m)^2 - p^m]} = 1 + p^m + 2p^{2m} + 2p^{3m} + 3p^{4m} + 3p^{5m} + \cdots$$

To take a more specific example, when $X = \mathrm{Gr}_2(\mathbb{A}^6)$ one has

$$N_m(X) = 1 + p^m + 2p^{2m} + 2p^{3m} + 3p^{4m} + 2p^{5m} + 2p^{6m} + p^{7m} + p^{8m}.$$

Now, the above examples are extremely trivial—for reasons we will explain below—but we can still use them to demonstrate the general idea of the Weil conjectures. Recall that the rational singular cohomology groups of the space $\mathbb{C}P^d$ (with its classical topology) are given by

$$H^i(\mathbb{C}P^d; \mathbb{Q}) = \begin{cases} \mathbb{Q} & \text{if } i \text{ is even and } 0 \leq i \leq 2d, \\ 0 & \text{otherwise.} \end{cases}$$

Likewise, the odd-dimensional cohomology groups of $\mathrm{Gr}_2(\mathbb{C}^6)$ all vanish and the even-dimensional ones are given by

i	0	1	2	3	4	5	6	7	8
$H^{2i}(X; \mathbb{Q})$	$\mathbb{Q}$	$\mathbb{Q}$	$\mathbb{Q}^2$	$\mathbb{Q}^2$	$\mathbb{Q}^3$	$\mathbb{Q}^2$	$\mathbb{Q}^2$	$\mathbb{Q}$	$\mathbb{Q}$

Note that in both examples the rank of $H^{2i}(X; \mathbb{Q})$ coincides with the coefficient of p^{im} in the formula for $N_m(X)$. This is the kind of phenomenon predicted by the Weil conjectures: relations between a formula for $N_m(X)$ and topological invariants of the corresponding complex algebraic variety.

In the cases of $\mathbb{P}^d$ and $\mathrm{Gr}_2(\mathbb{A}^d)$ (as well as all other Grassmannians), there is a very easy explanation for this coincidence. Consider the sequence of subvarieties

$$\emptyset \subseteq \mathbb{P}^0 \subseteq \mathbb{P}^1 \subseteq \cdots \subseteq \mathbb{P}^{d-1} \subseteq \mathbb{P}^d$$

Each complement $\mathbb{P}^i - \mathbb{P}^{i-1}$ is isomorphic to $\mathbb{A}^i$, and we can calculate the points of $\mathbb{P}^d$ by counting the points in all the complements and adding them up. As the number of points in $\mathbb{A}^i$ defined over $\mathbb{F}_{p^m}$ is just $(p^m)^i$, this immediately gives

$$N_m(\mathbb{P}^d) = 1 + p^m + p^{2m} + \cdots + p^{dm}$$

just as we found earlier.

But the same sequence of subvarieties—now considered over the complex numbers—gives a cellular filtration of $\mathbb{C}P^d$ in which there is one cell in every even dimension. The cells are the complements $\mathbb{C}P^i - \mathbb{C}P^{i-1}$. Of course this filtration is precisely what let's one calculate $H^*(\mathbb{C}P^d;\mathbb{Q})$.

The same kind of argument applies to Grassmannians, as well as other flag varieties. They have so-called "algebraic cell decompositions" given by the Schubert varieties, where the complements are disjoint unions of affine spaces. Counting the Schubert cells determines both $N_m(\mathrm{Gr}_k(\mathbb{C}^n))$ and $H^*(\mathrm{Gr}_k(\mathbb{C}^n);\mathbb{Q})$.

1.5. Deeper examples. If all varieties had algebraic cell decompositions then the Weil conjectures would be very trivial. But this is far from the case. In fact, only a few varieties have such decompositions. So we now turn to a more difficult example.

EXAMPLE 1.6. $X = V(x^3 + y^3 + z^3)$. This is an elliptic curve in $\mathbb{P}^2$ (recall that the genus of a degree D curve in $\mathbb{P}^2$ is given by $\binom{D-1}{2})$). So if we are working over $\mathbb{C}$, then topologically we are looking at a torus.

Counting the number of points of X defined over $\mathbb{F}_{p^m}$ is a little tricky. Weil gave a method for doing this in his original paper on the conjectures [**W5**], using some nontrivial results about Gauss and Jacobi sums. We will give this computation in Appendix B, but for now we'll just quote the results. When $p = 7$ (to take a specific case), computer calculations show that

$$N_1 = 9, \quad N_2 = 63, \quad N_3 = 324, \quad \text{and} \quad N_4 = 2331.$$

Weil's method gives the formula

$$N_m = 1 - \left[\left(\frac{-1 + 3\sqrt{3}i}{2} \right)^m + \left(\frac{-1 - 3\sqrt{3}i}{2} \right)^m \right] + 7^m,$$

which is in complete agreement. For convenience let $\alpha_1 = (-1 + 3\sqrt{3}i)/2$ and $\alpha_2 = \overline{\alpha}_1$.

Let's compare the above formula for N_m to the singular cohomology of the torus. It probably seems unlikely that the latter would ever let us predict the strange numbers α_1 and α_2! Despite this, there are several empirical observations we can make. If T is the torus, recall that $H^0(T;\mathbb{Q}) = H^2(T;\mathbb{Q}) = \mathbb{Q}$ and $H^1(T;\mathbb{Q}) = \mathbb{Q} \oplus \mathbb{Q}$. We can surmise that the even degree groups correspond to the 1 and 7^m terms, just as we saw for projective spaces and Grassmannians. The two $\mathbb{Q}$'s in $H^1(T;\mathbb{Q})$ are somehow responsible for the α_1^m and α_2^m terms. Note that $|\alpha_1| = |\alpha_2| = \sqrt{7}$, so this suggests that in general $H^j(X;\mathbb{Q})$ should contribute terms of norm $(7^{\frac{j}{2}})^m$ to the formula for $N_m(X)$. The way we have written the above formula further suggests that terms coming from $H^j(X;\mathbb{Q})$ are counted as negative when j is odd, but positive when j is even.

Also notice that $\alpha_1\alpha_2 = 7$. This should be compared to what one knows about $H^*(T;\mathbb{Q})$, namely that the product of two generators in H^1 gives a generator for H^2. This is related to Poincaré duality, and perhaps that is a better way to phrase this observation. The role of Poincaré duality is most evident in the $\mathrm{Gr}_2(\mathbb{A}^6)$ example done earlier, where one clearly sees it appearing as a symmetry in the formula for N_m. We can write this symmetry as follows. If d is the dimension of the variety X, then

$$\frac{N_m(X)}{p^{dm}} = N_{-m}(X)$$

where the right-hand-side means to formally substitute $-m$ for m in the formula for $N_m(X)$. The relation $\alpha_1\alpha_2 = 7$ says precisely that this equation is satisfied in the case of our elliptic curve.

Let's take a moment and summarize the observations we've made so far. Suppose X is a projective algebraic variety of dimension d defined by equations with integral coefficients. Let's also assume it's smooth, although the necessity of that assumption is not yet clear. Fix a prime p. We speculate that there is a formula

$$\begin{aligned} N_m(X) = 1 - [\alpha_{1,1}^m + \alpha_{1,2}^m + \cdots + \alpha_{1,b_1}^m] + [\alpha_{2,1}^m + \cdots + \alpha_{2,b_2}^m] - \cdots \\ + (-1)^{2d-1}[\alpha_{2d-1,1}^m + \cdots + \alpha_{2d-1,b_{2d-1}}^m] + p^{md} \end{aligned}$$

in which b_j is the rank of $H^j(X_{\mathbb{C}};\mathbb{Q})$ and $|\alpha_{j,s}| = p^{\frac{j}{2}}$. Note that $b_j = b_{2d-j}$, by Poincaré duality for $X_{\mathbb{C}}$ (as $X_{\mathbb{C}}$ is a $2d$-dimensional real manifold). We speculate that there is an associated duality between the coefficients $\alpha_{j,s}$ and $\alpha_{2d-j,s}$ which can be described either by saying that the set $\{\alpha_{j,s}\}_s$ coincides with the set $\{p^d/\alpha_{2d-j,s}\}_s$, or by the equality of formal expressions

$$\frac{N_m(X)}{p^{dm}} = N_{-m}(X).$$

We have just stated the Weil conjectures, although in a slightly rough form—we have, after all, not been so careful about what hypotheses on X are actually necessary. More formal statements will be given in the next section. For the moment we wish to explore a bit more, continuing our empirical investigations.

Perhaps more should be said about the mysterious coefficients $\alpha_{j,s}$. In the example of $X = V(x^3+y^3+z^3)$ and $p = 7$, α_1 and α_2 are algebraic integers—roots of the polynomial $x^2 + x + 7$. Could this polynomial have been predicted by the cohomology of the torus? Let's look at the variety $Y = V(y^2z - x^3 - xz^2)$, which is another elliptic curve in $\mathbb{P}^2$. When $p = 7$ computer calculations (see Appendix A) give that

$$N_1 = 8, \quad N_2 = 64, \quad N_3 = 344, \quad \text{and} \quad N_4 = 2304.$$

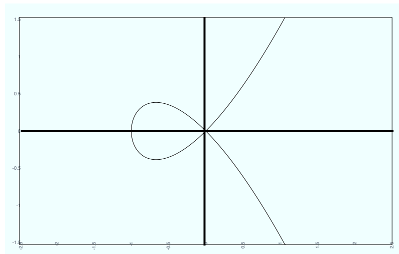
One can check that this agrees with the formula

$$N_m(Y) = 1 - \Big[(\sqrt{7}i)^m + (-\sqrt{7}i)^m\Big] + 7^m.$$

In this case the α_1 and α_2 are the roots of the polynomial $x^2 - 7$, which differs from our earlier example. So the moral is that the algebraic topology of the torus, while accounting for the overall *form* of a formula for N_m, does not determine the formula completely.

So far we have been working only with smooth, projective varieties. In the next two examples we explore whether these hypotheses are really necessary.

EXAMPLE 1.7 (Singular varieties). Again take $p = 7$, and let X be the nodal cubic in $\mathbb{P}^2$ given by the equation $y^2z = x^3 + x^2z$. This is the compactification—by adding the single point $[0 : 1 : 0]$—of the plane curve $y^2 = x^3 + x^2$ shown below:



Over the complex numbers, X is the quotient of a torus by one of its fundamental circles (or equivalently, X is obtained from S^2 by gluing two points together). So the cohomology groups are equal to $\mathbb{Z}$ in dimensions 0, 1, and 2. Based on our earlier examples, we might expect a formula $N_m = 1 - A^m + 7^m$ where $|A| = \sqrt{7}$. Yet simple computer calculations, explained in Appendix A below, show that

$$N_1 = 7, \quad N_2 = 49, \quad \text{and} \quad N_3 = 343.$$

The only value of A which is consistent with these numbers is $A = 1$, and of course this does not have the correct norm. So the Weil conjectures do not seem to hold for singular varieties.

This example can be better understood by blowing up the singular point of X. This blow-up $\tilde{X}$ turns out to be isomorphic to $\mathbb{P}^1$, and the map $\tilde{X} \to X$ just glues two points together to make the singularity (this is the easiest way to understand the topology of X over $\mathbb{C}$). It is then clear that one has

$$N_m(X) = N_m(\mathbb{P}^1) - 1 = [p^m + 1] - 1 = p^m$$

(for any base field $\mathbb{F}_p$). It is suggestive that we still have the formula

$$N_m = 1^m - A^m + p^m,$$

except that the norm of A is 1 rather than $p^{\frac{1}{2}}$. To fit this into context, consider X as the quotient S^2/A where $A = \{N, S\}$ consists of the north and south pole. Then the long exact sequence in cohomology gives

$$0 = \tilde{H}^0(S^2) \to \tilde{H}^0(A) \to H^1(X) \to H^1(S^2) = 0.$$

This gives $H^1(X) \cong \mathbb{Z}$, but what is important is that the $\mathbb{Z}$ in some sense 'came from' an H^0 group; this seems to be responsible for it it contributing terms to the formula for N_m of norm 1 rather than norm $p^{\frac{1}{2}}$.

What we are seeing here is the beginning of a long story, which would eventually take us to motives, mixed Hodge structures, and other mysteries. We will not pursue this any further at the moment, however. Suffice it to say that the Weil conjectures do not hold, as stated, for singular varieties, but that there may be some way of fixing them up so that they *do* hold.

EXAMPLE 1.8 (Affine varieties). Consider $X = \mathbb{A}^k - 0$. Then over the complex numbers this is homotopy equivalent to S^{2k-1}, hence its cohomology groups have a $\mathbb{Z}$ in dimension 0 and $2k-1$. The Weil conjectures might lead one to expect a formula $N_m = 1 - A^m$ where $|A| = q^{\frac{2k-1}{2}}$. What is actually true, however, is

$$N_m = (q^m)^k - 1 = (q^k)^m - 1^m.$$

So we find that the Weil conjectures—in the form we have given them—do not hold for varieties which are not projective.

This discrepancy can again be corrected with the right perspective. The key observation is that here one should not be looking at the usual cohomology groups, but rather at the *cohomology groups with compact support*. We will discuss this more in Chapter 2, but for now suffice it to say that these are just the reduced cohomology groups of the one-point compactification. In our case, the one-point compactification of $\mathbb{C}^k - 0$ is S^{2k} with the north and south poles identified. The cohomology with compact supports therefore has a $\mathbb{Z}$ in degrees 1 and $2k$, with the $\mathbb{Z}$ in degree one in some sense 'coming from' an H^0 as we saw in the previous example. Thus, the formula $N_m = -1^m + (q^k)^m$ now fits quite nicely with the Weil conjectures.

For the rest of this chapter we will continue to focus on smooth, projective varieties. But it is useful to keep the above two examples in mind, and to realize that with the right perspective some form of the Weil conjectures might work in a more general setting.

2. Formal statement of the conjectures

We will state the Weil conjectures in two equivalent forms. The first is very concrete, directly generalizing the discussion from the last section. The second approach, more common in the literature, uses the formalism of generating functions.

First we review some basic material. If K is a number field (i.e., a finite extension of $\mathbb{Q}$), recall that the ring of integers in K is the set $\mathcal{O}_K \subseteq K$ consisting of elements which satisfy a monic polynomial equation with integral coefficients. If $\wp \subseteq \mathcal{O}_K$ is a prime ideal, then $\mathcal{O}_K/\wp$ is a finite field.

In the last section we started with homogeneous polynomials $f_i \in \mathbb{Z}[x_0, \ldots, x_n]$ and considered their sets of zeros over extension fields of $\mathbb{F}_p$. One could just as well start with $f_i \in \mathcal{O}_K[x_0, \ldots, x_n]$ and look at solutions in extension fields of $\mathcal{O}_K/\wp$, for any fixed prime $\wp \subseteq \mathcal{O}_K$.

Let X be a variety defined over a finite field $\mathbb{F}_q$. One says that X **lifts to characteristic zero** if there is an algebraic variety $\mathfrak{X}$ defined over the ring of integers $\mathcal{O}$ in some number field, together with a prime ideal $\wp \subseteq \mathcal{O}$, such that $\mathcal{O}/\wp \cong \mathbb{F}_q$ and X is isomorphic to the mod $\wp$ reduction of $\mathfrak{X}$.

2.1. First form of the Weil conjectures. Let X be a smooth, projective variety defined over a finite field $\mathbb{F}_q$ (here $q = p^e$ for some prime p). Write $N_m(X) = \#X(\mathbb{F}_{q^m})$, and let d be the dimension of X.

Conjecture 2.2 (Weil conjectures, version 1).

(i) There exist non-negative integers $b_0, b_1, \ldots, b_{2d}$ and complex numbers $\alpha_{j,s}$ for $0 \le j \le 2d$ and $1 \le s \le b_j$ such that

$$N_m(X) = \sum_{j=0}^{2d} (-1)^j \left(\sum_{t=1}^{b_j} \alpha_{j,s}^m \right)$$

for all $m \geq 1$*. Moreover,* $b_0 = b_{2d} = 1$*,* $\alpha_{0,1} = 1$*, and* $\alpha_{2d,1} = q^d$*.*

(ii) The $\alpha_{j,s}$ *are algebraic integers satisfying* $|\alpha_{j,s}| = q^{j/2}$*.*

(iii) One has $b_j = b_{2d-j}$ *for all* j*, and the two sequences* $(\alpha_{j,1}, \ldots, \alpha_{j,b_j})$ *and* $(q^d/\alpha_{2d-j,1}, \ldots, q^d/\alpha_{2d-j,b_j})$ *are the same up to a permutation.*

(iv) Suppose that X *lifts to a smooth projective variety* $\mathfrak{X}$ *defined over the ring of integers* $\mathcal{O}$ *in a number field. Let* $\mathfrak{X}(\mathbb{C})$ *be the topological space of complex-valued points of* X*. Then* b_j *equals the* j*th Betti number of* $\mathfrak{X}(\mathbb{C})$*, in the sense of algebraic topology; that is,* $b_j = \dim_{\mathbb{Q}} H^j(\mathfrak{X}(\mathbb{C}); \mathbb{Q})$*.*

2.3. Second form of the conjectures. The equation given in 2.2(i) is somewhat awkward to work with, and it's form can be simplified by using generating functions. To see how, notice that if $N_m = A^m - B^m$ then one has an equality of formal power series

$$\sum_{m=1}^{\infty} N_m \frac{t^m}{m} = \log\Big(\frac{1}{1-At}\Big) - \log\Big(\frac{1}{1-Bt}\Big) = \log\Big(\frac{1-Bt}{1-At}\Big).$$

Generalizing, the equation in 2.2(i) says that

$$\sum_{m=1}^{\infty} N_m \frac{t^m}{m} = \log\Big(\frac{P_1(t)P_3(t)\cdots P_{2d-1}(t)}{P_0(t)P_2(t)\cdots P_{2d}(t)}\Big)$$

where $P_j(t) = \prod_s (1 - \alpha_{j,s}\, t)$.

It is traditional to define a formal power series

$$Z(X,t) = \exp\Big(\sum_{m=1}^{\infty} N_m \frac{t^m}{m}\Big).$$

This is called the **zeta function** of X. Using this, we may rephrase the Weil conjectures as follows:

Conjecture 2.4 (Weil conjectures, version 2).

(i) There exists polynomials $P_0(t), \ldots, P_{2d}(t)$ *such that*

$$Z(X,t) = \frac{P_1(t)P_3(t)\cdots P_{2d-1}(t)}{P_0(t)P_2(t)\cdots P_{2d}(t)}.$$

Moreover, $P_0(t) = 1 - t$ *and* $P_{2d}(t) = 1 - q^d t$*.*

(ii) The reciprocal roots of $P_j(t)$ *are algebraic integers whose norm is* $q^{j/2}$*.*

(iii) If $e = \sum_j (-1)^j \deg P_j(t)$*, then there is an identity of formal power series*

$$Z\Big(X, \frac{1}{q^d t}\Big) = (-1)^{b_d + a} q^{de/2} \cdot t^e \cdot Z(X,t).$$

where $b_d = \deg P_d(t)$ *and* a *is the multiplicity of* $-q^{-d/2}$ *as a root of* $P_d(t)$*.*

(iv) Suppose that X *lifts to a smooth projective variety* $\mathfrak{X}$ *defined over the ring of integers in a number field. Then* $\deg P_j(t)$ *coincides with the* j*th Betti number of* $\mathfrak{X}(\mathbb{C})$*, in which case the number* e *from (iii) is the Euler characteristic of* $\mathfrak{X}(\mathbb{C})$*.*

The statement in (i) is usually referred to as the *rationality of the zeta-function*. The statement in (ii) that the inverse roots of $P_j(t)$ have norm $q^{j/2}$ is called the *Riemann hypothesis* (for algebraic varieties over finite fields). The equation in (iii) is called the *functional equation* for $Z(X,t)$. These last two terms come from analogies with the classical Riemann zeta function which will be explained in the next section.

REMARK 2.5. We have been somewhat vague in specifying where the coefficients of the $P_i(t)$'s actually live. *A priori* they need only live in $\mathbb{C}$, but conjecture (2.2)(ii) immediately implies that the coefficients of the $P_i(t)$'s will actually be algebraic integers. We will see later that even more is true, and in fact the $P_i(t)$'s will all live in $\mathbb{Z}[t]$.

We'll briefly indicate the derivation of 2.4(iii), the other parts being obvious. From the statement in (2.2iii) we have

$$
\begin{aligned}
(2.6)\quad P_{2d-j}(t) = \prod_s (1-\alpha_{2d-j,s}t) &= \prod_s \Big(1-\frac{q^d}{\alpha_{j,s}}t\Big) \\
&= \Big(\prod_s \alpha_{j,s}\Big)^{-1}\cdot \prod_s (\alpha_{j,s}-q^d t) \\
&= (-1)^{b_d}\cdot (q^d t)^{b_d}\cdot \Big(\prod_s \alpha_{j,s}\Big)^{-1}\cdot \prod_s\Big(1-\frac{\alpha_{j,s}}{q^d t}\Big) \\
&= (-1)^{b_d}\cdot (q^d t)^{b_d}\cdot \Big(\prod_s \alpha_{j,s}\Big)^{-1}\cdot P_j\Big(\frac{1}{q^d t}\Big).
\end{aligned}
$$

Using that $b_j = b_{2d-j}$ and $\prod_s \alpha_{j,s}\cdot\prod_s \alpha_{2d-j,s} = (q^d)^{b_j}$ (which follows from (2.2iii)), we get

$$
\begin{aligned}
P_j(t)P_{2d-j}(t) &= (q^d t)^{2b_j}\cdot (q^d)^{-b_j}\cdot P_j\Big(\frac{1}{q_d t}\Big)\cdot P_{2d-j}\Big(\frac{1}{q_d t}\Big) \\
&= (q^d)^{\frac{b_j+b_{2d-j}}{2}}\cdot t^{(b_j+b_{2d-j})}\cdot P_j\Big(\frac{1}{q_d t}\Big)\cdot P_{2d-j}\Big(\frac{1}{q_d t}\Big).
\end{aligned}
$$

We may substitute this formula into the rational expression from (2.4i) and thereby replace all the products $P_j(t)P_{2d-j}(t)$, but the middle term $P_d(t)$ is left over. For this term one must use (2.6) itself, which says that

$$
P_d(t) = (-1)^{b_d}\cdot (q^d t)^{b_d}\cdot \pm q^{-\frac{b_d}{2}}\cdot P_d\Big(\frac{1}{q_d t}\Big) = \pm(-1)^{b_d}\cdot (q^d)^{\frac{b_d}{2}}\cdot t^{b_d}\cdot P_d\Big(\frac{1}{q_d t}\Big).
$$

Here we have used (2.2iii) to analyze the product $\prod_s \alpha_{d,s}$. We certainly have that $(\prod_s \alpha_{d,s})\cdot(\prod_s \alpha_{d,s}) = (q^d)^{b_d}$, and so $\prod_s \alpha_{d,s} = \pm (q^d)^{b_d/2}$. We must determine the sign. Every term $\alpha_{d,s}$ has a 'dual' term giving a product of q^d, so as long as a term is not its own dual its sign will cancel out of the product $\prod_s \alpha_{d,s}$. Some terms may be their own dual, however. This can only happen if the term is $q^{d/2}$ or $-q^{d/2}$. The terms $q^{d/2}$ are positive and therefore do not affect the sign of $\prod_s \alpha_{d,s}$. So the sign of this product is $(-1)^a$, where a is the number of terms $\alpha_{d,s}$ which are equal to $-q^{d/2}$.

Putting everything together, we have

$$
Z(X,t) = (-1)^{b_d+a}(q^d)^{-e/2}\cdot t^{-e}\cdot Z(X,\frac{1}{q^d t})
$$

and this is equivalent to the functional equation from (2.4iii).

3. Zeta functions

In this section we will take a brief detour and discuss the relation between various kinds of zeta functions—in particular, those of the Weil conjectures and the classical Riemann zeta function. This will lead us to a third form of the Weil conjectures, and will make it clear why the norm condition in (2.4ii) is called the *Riemann hypothesis*.

3.1. Riemann's function and its progeny. Recall that the Riemann zeta function is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \Big(1 - \frac{1}{p^s}\Big)^{-1}.$$

This is convergent and analytic in the range $Re(s) > 1$, but can be analytically continued to give a meromorphic function on the whole plane. This meromorphic function has zeros at all negative even integers (called the 'trivial' zeros), and these are the only zeros in the range $Re(s) < 0$. There are no zeros in the range $Re(s) > 1$, and the Riemann Hypothesis is that the only zeros in the so-called 'critical strip' $0 \leq Re(s) \leq 1$ are on the line $Re(s) = \frac{1}{2}$. The only pole of $\zeta(s)$ is a simple pole at $s = 1$.

It is useful to define a 'completed' version of the Riemann zeta function by

$$\hat{\zeta}(s) = \pi^{-\frac{s}{2}}\Gamma\Big(\frac{s}{2}\Big)\zeta(s).$$

Here Γ is the classical gamma-function of complex analysis. One is supposed to think of the above formula as adding an extra factor to the product $\prod_p (1-p^{-s})^{-1}$ corresponding to the 'prime at infinity'. It has the effect of removing the zeros at the even negative numbers, and adding a pole at $s = 0$. The Riemann Hypothesis is equivalent to the statement that all the zeros of $\hat{\zeta}(s)$ lie on the line $Re(s) = \frac{1}{2}$. Finally, we remark that $\hat{\zeta}(s)$ satisfies the so-called **functional equation** $\hat{\zeta}(s) = \hat{\zeta}(1-s)$.

For all of the above facts one may consult [**A**, Chapter 5.4], or any other basic text concerning the Riemann zeta function.

Let K be a number field with ring of integers $\mathcal{O}$. One may generalize the Riemann zeta function by defining

$$\zeta_K(s) = \sum_{n=1}^{\infty} \frac{\alpha_n}{n^s}$$

where α_n is the number of ideals $I \subseteq \mathcal{O}$ such that $\mathcal{O}/I$ has n elements (this is known to be finite). Then ζ_K is called the **Dedekind zeta function** for K, and $\zeta_{\mathbb{Q}}$ is just the classical Riemann zeta function. It is known that ζ_K is analytic in the range $Re(s) > 1$, and that it can be analytically continued to give a meromorphic function on the plane with a single, simple pole at $s = 1$. There is a product formula, namely

$$\zeta_K(s) = \prod_{\wp \subseteq \mathcal{O} \text{ prime}} \Big(1 - N(\wp)^{-s}\Big)^{-1}$$

where $N(\wp)$ denotes the order of the residue field $\mathcal{O}/\wp$.

One again has a completed version of this zeta function, here defined as

$$\hat{\zeta}_K(s) = D^s \left(\frac{\Gamma(s/2)}{\pi^{s/2}}\right)^{r_1} \left(\frac{\Gamma(s)}{(2\pi)^s}\right)^{r_2} \zeta_K(s)$$

where r_1 and r_2 are the numbers of real and complex places of K, and D is a certain invariant of K (the details are not important for us, but one may consult [**Lo**, Chapter VIII.2]). This completed zeta function again satisfies a functional equation $\hat{\zeta}_K(s) = \hat{\zeta}_K(1-s)$, and the generalized Riemann Hypothesis is the conjecture that all the zeros of $\hat{\zeta}_K$ lie on the line $Re(s) = \frac{1}{2}$.

Actually, we can generalize still further. Let X be a scheme of finite type over $\operatorname{Spec}\mathbb{Z}$. For every closed point $x \in X$, the residue field $\kappa(x)$ is a finite field (in fact, $\kappa(x)$ being a finite field is *equivalent* to x being a closed point in X). Write X_{max} for the set of closed points in X. Note that when $X = \operatorname{Spec} R$ this is just the set of maximal ideals in R.

When $x \in X_{max}$, define $Nx = \#\kappa(x)$, the number of elements in $\kappa(x)$. Then one defines

$$\zeta_X(s) = \prod_{x \in X_{max}} \Bigl(1 - (Nx)^{-s}\Bigr)^{-1}$$

in strict analogy with the classical Riemann zeta function. Note that when $X = \operatorname{Spec}\mathcal{O}$, where $\mathcal{O}$ is a ring of integers in a number field, this definition does indeed reduce to the Dedekind zeta function from above.

One must, of course, worry about whether the infinite product in the definition of ζ_X actually makes sense. One can show that the product converges absolutely when $Re(s) > \dim X$, but not much is known beyond this. It is conjectured that ζ_X has an analytic continuation to the entire plane, but this is only known in some special cases. We refer the reader to [**Se2**] for an introduction.

3.2. Schemes over finite fields. The function ζ_X simplifies in the special case where X is finite type over a finite field $\mathbb{F}_q$. The residue fields of closed points $x \in X$ will all be finite extensions of $\mathbb{F}_q$, and so one always has $Nx = q^{\deg(x)}$ where

$$\deg(x) = [\kappa(x) : \mathbb{F}_q].$$

For a general scheme X over $\operatorname{Spec}\mathbb{Z}$ there will be different bases for the exponentials in Nx as x varies, but for schemes over $\mathbb{F}_q$ this base is always just q. Based on this observation, it is reasonable to perform the change of variable $t = q^{-s}$ and write ζ_X as a function of t:

$$\zeta_X(s) = \prod_{x \in X_{max}} \Bigl(1 - t^{\deg(x)}\Bigr)^{-1}. \tag{3.4}$$

We claim that the expression on the right is none other than $Z(X, t)$. Incidentally, once we show this we will also have that $Z(X,t) \in \mathbb{Z}[[t]]$, as the above product certainly is a power series with integer coefficients.

The coefficient of t^n in (3.4) is readily seen to be

$$\#\Bigl\{x \in X_{max} : \deg(x) = n\Bigr\} + \frac{1}{2}\cdot\#\Bigl\{x \in X_{max} : \deg(x) = \frac{n}{2}\Bigr\} \\ + \frac{1}{3}\cdot\#\Bigl\{x \in X_{max} : \deg(x) = \frac{n}{3}\Bigr\} + \cdots$$

We have to relate this sum to $\#X(\mathbb{F}_{q^n})$.

If F is a field over $\mathbb{F}_q$, recall that an F-valued point of X is a map of $\mathbb{F}_q$-schemes $\operatorname{Spec} F \to X$. Specifying such a map is equivalent to giving a closed point $x \in X$ together with an $\mathbb{F}_q$-linear map of fields $\kappa(x) \to F$. It follows that

$$\#X(\mathbb{F}_{q^n}) = \sum_{j=0}^{\infty}\Big(\#\{x \in X_{max} : \deg(x) = j\} \cdot \#\operatorname{Hom}(\mathbb{F}_{q^j}, \mathbb{F}_{q^n})\Big).$$

But there are field homomorphisms $\mathbb{F}_{q^j} \to \mathbb{F}_{q^n}$ only when $j|n$, and the number of such homomorphisms which are $\mathbb{F}_q$-linear is just $\#\mathrm{Gal}(\mathbb{F}_{q^j}/\mathbb{F}_q) = j$. So we have

$$\begin{aligned}\#X(\mathbb{F}_{q^n}) &= \sum_{j|n}\Big(j \cdot \#\{x \in X_{max} : \deg(x) = j\}\Big)\\ &= n \cdot (\text{coefficient of } t^n \text{ in } (3.4)).\end{aligned}$$

We have therefore identified the product in (3.4) with $Z(X,t)$. That is to say, one has

$$\zeta_X(s) = Z(X, q^{-s}).$$

3.5. Zeta functions and the Weil conjectures.

Now we restrict to the case where X is smooth and projective over $\mathbb{F}_q$, in which case the Weil conjectures may be reinterpreted as statements about $\zeta_X(s)$.

What are the properties we would like for $\zeta_X(s)$? In analogy with the classical case, we would certainly like it to be meromorphic on the entire plane. But in fact it is even nicer: according to the first Weil conjecture (2.4i), $Z(X, q^{-s})$ is a rational function in q^{-s}. So $\zeta_X(s)$ is not only meromorphic, it is actually rational when regarded in the right way.

The Riemann Hypothesis (2.4iii) says something about the zeros and poles of $\zeta_X(s)$. Specifically, it says that $\zeta_X(s) = 0$ only if $|q^{-s}| = q^{-\frac{j}{2}}$ for some odd integer j in the range $1 \le j \le 2d-1$, where $d = \dim X$. This is equivalent to the statement that $Re(s) = \frac{j}{2}$, for some j in this range. Likewise, (2.4iii) says that ζ_X has a pole at s only if $|q^{-s}| = q^{-\frac{j}{2}}$ for some even integer j in the range $0 \le j \le 2d$. So we have that the zeros of ζ_X satisfy $Re(s) \in \{\frac{1}{2}, \frac{3}{2}, \ldots, \frac{2d-1}{2}\}$ and the poles of ζ_X satisfy $Re(s) \in \{0, 1, 2, \ldots, d\}$. Moreover, the only pole satisfying $Re(s) = 0$ is $s = 0$ and the only pole satisfying $Re(s) = d$ is $s = d$.

Notice the relation with the classical Riemann Hypothesis for $\hat{\zeta}$, which is morally the case where X is a compactified version of $\operatorname{Spec}\mathbb{Z}$. Here $d = 1$, and so the statement is that the zeros of $\hat{\zeta}$ lie only on the line $Re(s) = \frac{1}{2}$, and the only poles of $\hat{\zeta}$ are 0 and 1.

Finally we turn to the functional equation. Re-writing the equation in (2.4iii) in terms of s, one immediately gets

$$\begin{aligned}\zeta_X(s) = Z(X, q^{-s}) &= (-1)^{b_d + a} \cdot q^{e(s-\frac{d}{2})} \cdot Z(X, q^{s-d})\\ &= (-1)^{b_d + a} \cdot q^{e(s-\frac{d}{2})} \cdot \zeta_X(d-s).\end{aligned}$$

Recall that b_d is the degree of $P_d(t)$ and a is the multiplicity of $-q^{-d/2}$ as a root of $P_d(t)$. Alternatively, a can be taken to be the order of vanishing of ζ_X at the point $s = \frac{d}{2} - \frac{\pi}{\ln q} i$ (since this gives the same sign).

Here is a summary of everything we've just said:

CONJECTURE 3.6 (Weil conjectures, version 3). *Let X be a smooth, projective variety of finite type over the field $\mathbb{F}_q$. Let $d = \dim X$.*

(i) The zeta function $\zeta_X(s)$ is a rational function of q^{-s}. It has simple poles at $s = 0$ and $s = d$.

(ii) The zeros and poles of ζ_X lie in the critical strip $0 \le Re(s) \le d$. All of the zeros lie on the lines $Re(s) = \frac{j}{2}$ where j is an odd integer in the range $1 \le j \le 2d - 1$. The poles lie on the lines $Re(s) = j$ where j is an integer in the range $0 \le j \le d$.

(iii) ζ_X satisfies a functional equation of the form $\zeta_X(s) = \pm q^{e(s-\frac{d}{2})}\zeta_X(d - s)$ for some integer e.

(iv) Suppose that X lifts to a smooth projective variety $\mathfrak{X}$ defined over the ring of integers in a number field. Then when j is odd, the number of zeros of ζ_X on the line $Re(s) = \frac{j}{2}$ coincides with the jth Betti number of $\mathfrak{X}(\mathbb{C})$. When j is even, the jth Betti number of $\mathfrak{X}(\mathbb{C})$ is the number of poles of ζ_X on the line $Re(s) = \frac{j}{2}$.

4. A plan to prove the conjectures

The Weil conjectures were introduced, quite briefly, in [**W5**]. Weil spent most of that paper working out a class of examples, stated his conjectures in the last pages, and then stopped without further remark. It is not until the ICM lecture [**W6**] that one finds a published suggestion for how one might go about proving them.

Let $X \hookrightarrow \mathbb{P}^n$ be a smooth, projective variety over $\mathbb{F}_q$. There is a canonical morphism $F\colon X \to X$ which is the identity on the underlying topological space of X and induces the qth power map $\mathcal{O}_X(U) \to \mathcal{O}_X(U)$ for every open set $U \subseteq X$. This is called the *geometric Frobenius morphism.* (Note that if $q = p^e$ then there is also a map of schemes $X \to X$ which induces the pth power—rather than the qth power—on the ring of functions, but this is not a morphism of schemes over $\mathbb{F}_q$).

For any extension field $\mathbb{F}_q \hookrightarrow E$, let $X(E)$ denote the set of maps $\operatorname{Spec} E \to X$ over $\operatorname{Spec} \mathbb{F}_q$. Then F induces the map

$$F\colon X(E) \to X(E), \qquad (x_0, \ldots, x_n) \mapsto (x_0^q, \ldots, x_n^q).$$

Let $\overline{X} = X \times_{\operatorname{Spec} \mathbb{F}_q} (\operatorname{Spec} \overline{\mathbb{F}}_q)$ be the base extension of X to $\overline{\mathbb{F}}_q$. There are three morphisms $\overline{X} \to \overline{X}$ which arise naturally. One is the map $F \times id$. Another is the map $\overline{X} \to \overline{X}$ which is the identity on topological spaces and is the qth power map on rings of functions; we'll call this map $F_{\overline{X}}$. Finally, there is a third map which can be defined as follows. Let $\sigma \in \operatorname{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ be the Frobenius element $\alpha \mapsto \alpha^q$. Recall that $\operatorname{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q) \cong \hat{\mathbb{Z}}$ and σ is a topological generator. Then one also has the map of schemes $id \times \sigma\colon \overline{X} \to \overline{X}$, called the *arithmetic Frobenius morphism.* Note that $\overline{F} = F \times \sigma = (F \times id) \circ (id \times \sigma)$.

The only one of these three maps $\overline{X} \to \overline{X}$ which is a map of schemes over $\overline{\mathbb{F}}_q$ is $F \times id$. Because of this, it is common to just write F as an abbreviation for $F \times id$. Be careful of the distinction between F and $F_{\overline{X}}$.

If $\overline{X}(\overline{\mathbb{F}}_q)$ denotes the set of maps $\operatorname{Spec} \overline{\mathbb{F}}_q \to X$ over $\operatorname{Spec} \overline{\mathbb{F}}_q$, then F induces the map

$$F\colon \overline{X}(\overline{\mathbb{F}}_q) \to \overline{X}(\overline{\mathbb{F}}_q), \qquad (x_0, \ldots, x_n) \mapsto (x_0^q, \ldots, x_n^q).$$

The fixed points of this map are therefore precisely the points of $X(\mathbb{F}_q)$, and more generally the fixed points of the mth power F^m are the points of $X(\mathbb{F}_{q^m})$.

With this point of view, the Weil conjectures become about understanding the number of fixed points of powers of F. In algebraic topology, the most basic tool one has for understanding fixed points is the Lefschetz trace formula. This says that if $f\colon Z \to Z$ is a continuous endomorphism of a compact manifold then the number of fixed points of f (counted with appropriate multiplicities) is the same as the Lefschetz number

$$\Lambda(f) = \sum_{j=0}^{\infty} (-1)^j \operatorname{tr}\Big[f^*\big|_{H^j(X;\mathbb{Q})}\Big].$$

Note that this is really a finite sum, of course.

4.1. Cohomological approach. Weil proposed that one might be able to attach to the scheme $\overline{X}$ a sequence of algebraically defined cohomology groups which we'll call $H^j_W(\overline{X})$. These should ideally be finite-dimensional vector spaces defined over some characteristic 0 field E, and should be non-vanishing only in the range $0 \le j \le 2d$, where $d = \dim X$. There should be a Lefschetz trace formula analagous to the one above. So one would have

$$N_m(X) = \#X(\mathbb{F}_{q^m}) = \#\{\text{fixed points of } F^m\} = \sum_{j=0}^{2d} (-1)^j \operatorname{tr}\Big[(F^*)^m\big|_{H^j_W(\overline{X})}\Big].$$

To explain how this helps with the conjectures, we need a simple lemma from linear algebra:

Lemma 4.2. *Let V be a finite-dimensional vector space over a field k, and let $L\colon V \to V$ be a linear transformation. Define $P_L(t) = \det(I - Lt) \in k[t]$. Then one has an identity of formal power series*

$$\log\left(\frac{1}{P_L(t)}\right) = \sum_{m=1}^{\infty} \operatorname{tr}(L^m)\cdot \frac{t^m}{m}.$$

Proof. We may as well extend the field, and so we can assume k is algebraically closed. Using Jordan normal form, we can write $L = D + N$ where D is represented by a diagonal matrix and N is strictly upper triangular. Then $P_L(t) = P_D(t)$ and $\operatorname{tr}(L^m) = \operatorname{tr}(D^m)$, hence one reduces to the case where $L = D$. But this case is obvious. $\square$

Now we simply compute:

$$\begin{aligned}
Z(X,t) &= \exp\Big(\sum_{m=1}^{\infty} N_m \frac{t^m}{m}\Big)\\
&= \exp\Big(\sum_{m=1}^{\infty}\sum_{j=0}^{2d}(-1)^j \operatorname{tr}\Big[(F^*)^m|_{H^j_W(\overline{X})}\Big]\cdot\frac{t^m}{m}\Big)\\
&= \prod_{j=0}^{2d}\Big(\exp\Big(\sum_{m=1}^{\infty}\operatorname{tr}\Big[(F^*)^m|_{H^j_W(\overline{X})}\Big]\cdot\frac{t^m}{m}\Big)\Big)^{(-1)^j}\\
&= \prod_{j=0}^{2d} P_j(t)^{(-1)^{j+1}} \qquad \text{by Lemma 4.2,}
\end{aligned}$$

where $P_j(t) = \det(I - \phi_j t)$ with $\phi_j = F^*|_{H^j_W(\overline{X})}$.

Notice that this gives the rationality of $Z(X,t)$, as predicted in (2.4i). The expected equality $P_0(t) = 1-t$ would follow from knowing $H^0_W(\overline{X})$ is one-dimensional and $F^* = id$ on this group (as would happen in algebraic topology). The conjecture $P_{2d}(t) = 1-q^d t$ likewise suggests that $H^{2d}_W(\overline{X})$ should be one-dimensional, with F^* acting as multiplication by q^d.

One can continue in this way, re-interpreting the Weil conjectures as expected properties of the cohomology theory H^*_W. For instance, note that the $\alpha_{j,s}$'s of (2.2i) will be the reciprocal roots of $P_j(t)$, which are just the eigenvalues of F^* acting on $H^j_W(\overline{X})$. This is so important that we will state it again:

(**) *The numbers $\alpha_{j,s}$ of (2.2i) are the eigenvalues of F^* acting on $H^j_W(\overline{X})$.*

We will next show that (2.2iii) is a consequence of a Poincaré Duality theorem for H^*_W. It is reasonable to expect a cup product on $H^*_W(\overline{X})$ making it into a graded ring, and for F^* to be a ring homomorphism. Poincaré Duality should say that when X is smooth and projective then

$$H^j_W(\overline{X}) \otimes H^{2d-j}_W(\overline{X}) \xrightarrow{\cup} H^{2d}_W(\overline{X})$$

is a perfect pairing, and hence $\dim_K H^j_W(\overline{X}) = \dim_K H^{2d-j}_W(\overline{X})$. If F^* acts on $H^{2d}_W(\overline{X})$ as multiplication by q^d, it follows immediately that the eigenvalues $\{\alpha_{j,s}\}$ (counted with multiplicity) of F^* acting on H^j_W are related to those of F^* acting on H^{2d-j}_W by the formula

$$\{q^d/\alpha_{j,s}\}_s = \{\alpha_{2d-j,s}\}_s.$$

But this is exactly what is required by (2.2iii), or the equivalent statement (2.4iii).

4.3. The Künneth theorem. Let X and Y be two smooth, projective varieties over $\mathbb{F}_q$. Then $(X\times Y)(\mathbb{F}_{q^m}) = X(\mathbb{F}_{q^m}) \times Y(\mathbb{F}_{q^m})$, and so $N_m(X\times Y) = N_m(X)\cdot N_m(Y)$. If we have formulas

$$N_m(X) = \sum_{j,s}(-1)^j \alpha_{j,s}^m \qquad \text{and} \qquad N_m(X) = \sum_{k,t}(-1)^k \beta_{k,t}^m$$

as specified by the Weil conjectures, multiplying them together gives a similar formula

$$N_m(X \times Y) = \sum_l (-1)^l \sum_{\substack{j+k=l \\ s,t}} (\alpha_{j,s}\beta_{k,t})^m.$$

In terms of our cohomological interpretation, this says that if we know the eigenvalues of F^* on $H^*_W(X)$ and on $H^*_W(Y)$, then their products give the eigenvalues of F^* on $H^*_W(X \times Y)$.

The cup product on $H^*_W(X \times Y)$ allows us to define a map of graded rings

$$\kappa\colon H^*_W(X) \otimes H^*_W(Y) \to H^*(X \times Y)$$

in the usual way: $\kappa(a \otimes b) = \pi_1^*(a) \cup \pi_2^*(b)$. The above observations about the eigenvalues of F^* are in exact agreement with the hypothesis that κ is an isomorphism. So it is reasonable to expect our conjectural theory H^*_W to satisfy the Künneth theorem.

4.4. Behavior under base-change. We will postpone a cohomological interpretation of (2.2ii) until Chapter 3, as this will require a detour through Hodge-Lefschetz theory. Let us instead move on to (2.2iv), the comparison with ordinary singular cohomology. For this, we need to move outside of the realm of finite fields.

Let us suppose that H^*_W can be defined for any scheme of reasonable type. In particular, it can be defined for schemes over $\mathbb{C}$. It is reasonable to expect a natural transformation

$$H^*_W(\mathfrak{X}) \to H^*_{\text{sing}}(\mathfrak{X}(\mathbb{C}); E)$$

of ring-valued functors, for $\mathbb{C}$-schemes $\mathfrak{X}$ (remember that E is the coefficient field of H^*_W). One can hope that when $\mathfrak{X}$ is smooth and projective this is an isomorphism.

Now suppose that $\mathfrak{X}$ is a scheme defined over the ring of integers $\mathcal{O}$ in a number field. Let $\wp \subseteq \mathcal{O}$ be a prime, and let $\mathfrak{X}_\wp$ be the pullback of $\mathfrak{X}$ along the map $\operatorname{Spec} \mathcal{O}_\wp \to \operatorname{Spec} \mathcal{O}$. One can choose an embedding $\mathcal{O}_\wp \hookrightarrow \mathbb{C}$, and of course one has the projection $\mathcal{O}_\wp \twoheadrightarrow \mathcal{O}_\wp/\wp$; note that $\mathcal{O}_\wp/\wp$ is a finite field. One forms the following diagram of pullbacks:

$$\begin{tikzcd}
X \arrow[r] \arrow[d] & \mathfrak{X}_\wp \arrow[d] & \mathfrak{X}_{\mathbb{C}} \arrow[l] \arrow[d] \\
\operatorname{Spec} \mathcal{O}_\wp/\wp \arrow[r] & \operatorname{Spec} \mathcal{O}_\wp & \operatorname{Spec} \mathbb{C}. \arrow[l]
\end{tikzcd}$$

We then have induced maps $H^*_W(\mathfrak{X}_\wp) \to H^*_W(X)$ and $H^*_W(\mathfrak{X}_\wp) \to H^*_W(\mathfrak{X}_{\mathbb{C}}) \to H^*(\mathfrak{X}(\mathbb{C}); E)$. The Weil conjecture of (2.2iv) will follow if one knows these induced maps are isomorphisms.

4.5. The coefficient field. At this point we have built up an impressive amount of speculation about this mysterious cohomology theory H^*_W. Does such a thing really exist? The first thing one is forced to consider is the choice of the coefficient field E.

Of course it seems reasonable, and desirable, to just have $E = \mathbb{Q}$. But an early observation due to Serre shows that with this coefficient field no such H^*_W can exist. In fact, no such cohomology theory exists in which E is a subfield of $\mathbb{R}$. The explanation is as follows.

Suppose that one has a theory H^*_W defined for schemes over a given field F, and let E be the coefficient field of the theory. For any F-scheme X, let $\operatorname{Hom}(X, X)$

denote the endomorphism monoid of X (the monoid of self-maps in the category of F-schemes). Since H_W^* is a functor, it follows that $\mathrm{Hom}(X,X)$ acts on $H_W^*(X)$.

Now suppose X is an abelian variety. This means there is a map $\mu\colon X\times X \to X$ which is commutative, associative, unital, and there is an additive inverse map $\iota\colon X \to X$. Let $\mathrm{End}(X)$ denote the set of homomorphisms $X \to X$ regarded now as a *ring*, where the multiplication is composition and the addition is induced by μ. Specifically, if $f,g\colon X \to X$ then $f+g$ is defined to be the composite

$$X \xrightarrow{\Delta} X\times X \xrightarrow{f\times g} X\times X \xrightarrow{\mu} X.$$

The monoid $\mathrm{Hom}(X,X)$ from the last paragraph is just the multiplicative monoid of R.

One can check that $H^1_W(X)$ is necessarily a module over $\mathrm{End}(X)$. This will not be true for the other $H^k_W(X)$'s, but works for H^1 because of the isomorphism $\pi_1^*\oplus\pi_2^*\colon H^1_W(X)\oplus H^1_W(X) \to H^1_W(X\times X)$ given by the Künneth theorem. See Exercise 4.6 at the end of this section.

When X is an elliptic curve, quite a bit is known about the endomorphism ring $\mathrm{End}(X)$. In particular, it is a characteristic zero integral domain of finite rank over $\mathbb{Z}$, and $\mathrm{End}(X)\otimes\mathbb{R}$ is either $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. See [**Si**, Cor. III.9.4]. Much more is known about $\mathrm{End}(X)$ than just this statement, but this is all that we will need. An elliptic curve is called *supersingular* precisely when $\mathrm{End}(X)\otimes\mathbb{R}\cong\mathbb{H}$.

If our speculation about H_W^* is correct, then for X an elliptic curve $H^1_W(X)$ must be a two-dimensional vector space over the coefficient field E. So $\mathrm{End}(X)$ has a representation on E^2. But if $E\subseteq\mathbb{R}$, one then obtains a representation of $\mathrm{End}(X)\otimes\mathbb{R}$ on $\mathbb{R}^2$ by extending the coefficients. This is impossible in the case where X is supersingular, as there is no representation of $\mathbb{H}$ on $\mathbb{R}^2$. So we have obtained a contradiction; there is no theory H_W^* having the expected properties and also having $E\subseteq\mathbb{R}$.

Exercise 4.6. Verify that $\mathrm{End}(X)$ is indeed a ring, with the addition and multiplication defined above. If $f,g\colon X\to X$, show that there is a commutative diagram

$$\begin{tikzcd}
 & H^1_W(X\times X) \arrow[r, "(f\times g)^*"] & H^1_W(X\times X) \arrow[dr, "\Delta^*"] & \\
H^1_W(X) \arrow[ur, "\mu^*"] \arrow[dr, "D"'] & & & H^1_W(X), \\
 & H^1_W(X)\oplus H^1_W(X) \arrow[uu, "\cong"', "\pi_1^*\oplus\pi_2^*"] \arrow[r, "f^*\oplus g^*"'] & H^1_W(X)\oplus H^1_W(X) \arrow[uu, "\pi_1^*\oplus\pi_2^*", "\cong"'] \arrow[ur, "\sigma"'] &
\end{tikzcd}$$

where $D(a)=(a,a)$ and $\sigma(a,b)=a+b$. Use this to verify that $H^1_W(X)$ is a module over $\mathrm{End}(X)$.

5. Some history of the proofs of the conjectures

Nice summaries of the work on the Weil conjectures can be found in [**Ka**] and [**M3**]. Here we will only give a very brief survey.

When Weil made his conjectures, he was generalizing what was already known for curves. In fact it was Weil himself who had proven the Riemann Hypothesis in this case, a few years earlier. The challenge was therefore was to prove the conjectures for higher dimensional varieties. The first to be proven in this generality was the rationality of the zeta function. This was done by Dwork [**Dw**], using an approach via p-adic analysis that was very different from what we outlined above. In particular, Dwork's approach is entirely non-cohomological.

Independently, Grothendieck, M. Artin, and others were developing étale cohomology. This work produced a *family* of cohomology theories H_W^*, one for each prime l different from the characteristic of the ground field. These so-called 'l-adic' cohomology theories had $\mathbb{Q}_l$ as their coefficient field.

Grothendieck and his collaborators proved the Lefschetz trace formula and Poincaré Duality for these l-adic cohomology theories, and in this way established (2.4i) and (2.4iii). They also proved the necessary comparison theorems to singular cohomology, from which (2.4iv) follows. All of this requires quite a bit of work and machinery.

Two things were left unanswered by this original work of Grothendieck et al. The first is the Riemann Hypothesis (2.4ii). The second is the so-called question of "independence of l". Each l-adic cohomology theory $H^*(-;\mathbb{Q}_l)$ gives rise to a Lefschetz trace formula and a resulting factorization

$$Z(X,t) = \prod_i [P_i(t)_l]^{(-1)^{i+1}}.$$

However, the polynomials $P_i(t)_l$ could only be said to lie in $\mathbb{Q}_l[t]$ rather than $\mathbb{Z}[t]$, and it was not clear whether different choices of l led to different polynomials.

Grothendieck and Bombieri independently developed a plan for answering these final questions. Everything was reduced to two conjectures on algebraic cycles which Grothendieck called the "Standard Conjectures". See [**G2**] and [**Kl1**]. These conjectures are very intriguing, and really explain the geometry underlying the Weil conjectures. But they have so far resisted all attempts on them, and remain open except in special cases.

The Riemann Hypothesis and the independence of l were proven for smooth, projective varieties by Deligne in the early 1970s. In the earlier papers [**D1**] and [**D2**] Deligne had proven the Riemann hypothesis for $K3$ surfaces and for certain complete intersections, but these results were eclipsed by the complete solution two years later in [**D4**]. Deligne's very ingenious method avoided the Standard Conjectures completely, much to everyone's surprise. For a very nice summary, see [**Ka**].

Closing thoughts

In this chapter we have given a quick overview of the Weil conjectures and how they inspired the search for a suitable cohomology theory for algebraic varieties. This is only the beginning of a long story with many branches, some of which we now outline.

(1) We have given cohomological interpretations for all aspects of the Weil conjectures except two. These are the Riemann hypothesis and the conjecture that the polynomials $P_i(t)$ (appearing in the rational expression for the zeta function) should have *integral* coefficients. Cohomological explanations for these two conjectures were provided by Grothendieck's "Standard Conjectures". These will be described in Chapter 3 below.
(2) The Riemann hypothesis for curves was proven by Weil in the 1940s. Later Grothendieck gave a proof using the Riemann-Roch theorem, and Stepanov gave an elementary proof. Weil's proof is very interesting, however, because he

was able to use the Jacobian variety of the curve as a geometric substitute for the cohomology group H^1. This idea of cohomology theory having a geometric "motive" underlying it was later developed by Grothendieck and led to his conjectural category of motives.

(3) Grothendieck, Artin, Verdier, and others developed étale cohomology. This required a vast amount of machinery, and has been very influential. We will describe étale cohomology in Chapter 4.

(4) Zeta functions are part of a much broader class of objects called *L-functions*. Grothendieck was able to use étale cohomology to generalize the Weil conjectures, and get information not only about the zeta functions of algebraic varieties over finite fields but also about a larger class of L-functions.

(5) Dwork proved the rationality of the zeta function using methods of p-adic analysis, and later he was able to prove most of the Weil conjectures for hypersurfaces using those techniques. This work then led to the development of p-adic cohomology theories for characteristic p varieties, building off of p-adic differential calculus. Monsky and Washnitzer developed a theory called *formal cohomology*, and Grothendieck outlined a theory—developed by Berthelot—called *crystalline cohomology*. In later years Berthelot also developed a theory called *rigid cohomology*, and this has been very influential as of late.

Appendices to Chapter 1

A. Computer calculations

In the course of learning any area of mathematics, it is nice to sit down and work out specific examples. As mathematics has become more sophisticated, however, working out examples has become harder and harder. Counting—by hand—the number of points of a variety defined over finite fields $\mathbb{F}_{p^k}$ is very unpleasant. But modern computers can help with this somewhat, and in this section we will describe some simple tools for getting started.

Now, let's be honest. Given a set of equations, the number of computations necessary to count solutions over $\mathbb{F}_{p^k}$ is going to grow exponentially with k. So even computers are going to be very limited in the number of examples they can actually work out. But being able to look at a few examples is better than not being able to look at any.

There are different computer packages available for handling arithmetic in finite fields. Mathematica can handle this, as can Macaulay2. Here we will describe how to do this using a software package called Sage, which is an extension of the Python programming language. Sage is open source software which is freely available for download, and Python is a very wonderful programming language—it is easy to use, intuitive, and its style works well for mathematicians.

To download sage, visit the website

`www.sagemath.org`

Sage can be run either from a "command line" or from a "notebook". For simplicity, we will assume it is being run from the command line. This will mean that the program gives the prompt "sage:" when it waits for input. One can type "2ˆ3+17*5", and after hitting return the software will evaluate that expression.

Try the following commands:

```
 sage:   E=GF(5)
 sage:   for a in E:
.... :       print a,a^2,a^3
.... :   [Return]
```

Note that we have written "`[Return]`" to indicate that the user should press the Return or Enter key. Also note that the indentation in the above code is important: Sage and Python use indentation in a structural way, to control looping and conditional statements. The identation in the above example tells Sage that the `print` command is part of the `for` loop.

Upon entering the above commands, Sage will output the following list showing the elements of $\mathbb{F}_5$ (called $GF(5)$ in Sage), as well as their squares and cubes:

```
0  0  0
1  1  1
2  4  3
3  4  2
4  1  4.
```

For something more sophisticated, try:

```
 sage:   F.<z>=GF(25)
 sage:   for a in F:
.... :        print a,"\t",a^2
.... :   [Return]
```

The "z" which appears in "`F.<z>`" is a variable name for a primitive element of this extension field of $GF(5)$. The `\t` in the `print` command produces a tabbed space between the outputs a and a^2.

Try some arithmetic in E and F:

```
sage:   (1+3*z)^3
sage:   2^4
sage:   E(2)^4
sage:   F(2)^4
```

Note that Sage interprets the number 2 in the second line as an ordinary integer. If we want to talk about 2 as an element of F, Sage requires us to use "`F(2)`". However, to refer to the element $2z \in F$ we can write either `2*z` or `F(2)*z`; Sage understands that they mean the same thing.

To find out whether $1 + 3z$ has a square root in F, we could do the following:

```
 sage:   for a in F:
.... :        if a^2==1+3*z:
.... :            print a," is a square root of 1+3z"
.... :   [Return]
```

You will note that Sage has no output upon running this routine—which just tells us that it didn't find any square roots. Try running a similar routine to find the cube roots of $1 + 3z$ in F.

One can define functions in Sage. Here is a simple example to try:

```
 sage:   def f(a,b):
.... :        return a^2+3*a*b
.... :   [Return]
 sage:   f(1+3*z,2+z)
```

Sage has various built-in functions for dealing with finite fields. The two we will need return the order of a field and the multiplicative order of a given element. Here are some samples:

```
sage:   order(F)
25
sage:   multiplicative_order(1+3*z)
8
```

Here is a short function which will return a generator for the multiplicative group of units of a given finite field. Note that Sage understands the idea of dummy variables, and so it knows that the "F" in the code below is not the F we have globally defined to be $GF(25)$.

```
 sage:   def mult_generator(F):
.... :        for a in F:
.... :              if a==0:
.... :                  continue
.... :              if order(F)-1==multiplicative_order(a):
.... :                  return a
.... :   [Return]
```

Now try the following two commands:

```
sage:   mult_generator(E)
sage:   mult_generator(F)
```

At this point we have all the techniques we need to have Sage count solutions to equations for us—nothing fancy, just brute force enumeration. The following function takes two inputs: a field F and a function of three variables f. It then returns the number of triples $(a, b, c) \in F^3$ such that $f(a, b, c) = 0$.

```
 sage:   def count3(F,f):
.... :        output=0
.... :        for a in F:
.... :             for b in F:
.... :                  for c in F:
.... :                       if f(a,b,c)==0:
.... :                            output=output+1
.... :        return output
.... :   [Return]
```

There are a couple of important observations to make about the above code. First, recall that Sage uses indentation for structural purposes. It is very important that the "`return output`" command have the same indentation as the "`for a in F:`" command. This tells Sage that the `return` command should be executed after the "`for a in F`" loop is completely finished.

Secondly, because it's easy to make mistakes when typing, it can be a pain to define routines like `count3` via Sage's command line. It is more convenient to use a text editor to put the code into a file, let's say one called `weil.sage`. The command

```
sage:   attach "weil.sage"
```

will then load the file into Sage's memory and execute all the commands.

To use the above counting routine, try:

```
 sage:   def f(x,y,z):
.... :        return x^3+y^3+z^3
.... :   [Return]
 sage:   F.<x>=GF(5^2)
 sage:   count3(F,f)
```

Sage should return the number 865, which is the number of (affine) solutions to the equation $x^3 + y^3 + z^3$ in $\mathbb{F}_{25}$. To get the number of projective solutions one of course subtracts 1 and divides by 24, to get 36.

One can use `count3` to count the number of solutions of other three-variable functions as well. For instance, try:

```
 sage:   def g(x,y,z):
.... :        return x^2+x*y^3-y*z
.... :   [Return]
 sage:   E.<w>=GF(5^3)
 sage:   count3(E,g)
```

As the size of the finite field gets large, it can take Sage a long time to do the above kind of brute force enumeration. It pays to use a little intelligence now and then. For instance, suppose we want to count the number of points of the projective variety defined by $x^3z^2 + xyz^3 - x^3y^2 = 0$. When $z = 0$ we get $x^3y^2 = 0$, which means either $x = 0$ or $y = 0$. So there are two solutions when $z = 0$, namely $[1 : 0 : 0]$ and $[0 : 1 : 0]$. When $z \neq 0$ we can normalize z to be 1, which means we are then interested in the *affine* solutions to $x^3 + xy - x^3y^2 = 0$. It is much faster for Sage to count solutions to this equation—and to add two to the answer—then to count the number of solutions to the original equation.

We close this section with one last example, which will be used in Appendix B. It serves to demonstrate Sage's syntax for complex arithmetic and list manipulation.

Let $F = \mathbb{F}_q$, and recall that the group of units F^* is cyclic. Let $\zeta = \frac{-1+\sqrt{3}i}{2}$. If $3|q-1$ then there is a group homomorphism $\chi\colon F \to \mathbb{C}$ such that $\chi(g) = \zeta$. By convention we set $\chi(0) = 0$. In Appendix B we will have to evaluate sums of the form

$$J(\chi) = \sum_{u_1+u_2+u_3=0} \chi(u_1)\chi(u_2)\chi(u_3)$$

where the u_i's range over all elements of F.

First note that Sage has built-in capabilities for complex arithmetic. Try

```
sage:   zeta=(-1+sqrt(3)*I)/2
sage:   zeta^2
sage:   zeta^3
sage:   expand(zeta^2)
```

You will note that Sage performs the operations algebraically, without any simplification, unless it is given the `expand` command.

The following routine takes a field F and a multiplicative generator g, and returns the sum $J(\chi)$. For some reason, the Sage routines for complex arithmetic are somewhat slow—in the sense that doing 100 computations takes a noticeable amount of time. The code avoids this issue by putting off all complex arithmetic until the end. Each term $\chi(u_1)\chi(u_2)\chi(u_3)$ is either 1, ζ, or $\overline{\zeta}$, and what the code does is count the number of times each possibility appears. Then only at the very end does it form the appropriate linear combination of complex numbers. Here is the code:

```
 sage:   zeta=(-1+sqrt(3)*I)/2
 sage:   zetabar=(-1-sqrt(3)*I)/2
 sage:   def J(g,F):
.... :      count=[0,0,0]
.... :      list=[]
.... :      k=1
.... :      while k<=order(F)-1:
.... :         list.append(g^k)
.... :         k=k+1
.... :      a=1
.... :      while a<=order(F)-1:
.... :         b=1
.... :         while b<=order(F)-1:
.... :            c=-g^a-g^b
.... :            if c==0:
.... :               b=b+1
.... :               continue
.... :            k=1
.... :            while k<=len(list):
.... :               if c==list[k-1]:
.... :                  break
.... :               k=k+1
.... :            exponent=mod(a+b+k,3)
.... :            count[exponent]=count[exponent]+1
.... :            b=b+1
.... :         a=a+1
.... :      output=count[0]+count[1]*zeta+count[2]*zetabar
.... :      return expand(output)
```

To use the code, try the following:

```
sage:  F=GF(31)
sage:  z=mult_generator(F)
sage:  J(z,F)
```

Enjoy playing around!

There are nice references for learning more about both Sage and Python. Tutorials and reference manuals can be found at the following two websites:

`www.sagemath.org/doc` and `www.python.org/doc`

B. Computations for diagonal hypersurfaces

In [**W5**] Weil verified his conjectures for hypersurfaces defined by an equation of the form $a_0x_0^d + a_2x_2^d + \cdots + a_kx_k^d = 0$. His technique involved writing a formula for N_m in terms of so-called Gauss and Jacobi sums, and then appealing to certain theorems from number theory. This section will describe Weil's method.

There are two main reasons we have included this material. Foremost, these hypersurfaces provide the first examples of the Weil conjectures which are not trivial in the way that projective spaces and Grassmannians are. The fact that these examples actually *work* really shows that there is something interesting going on. From a topological perspective, hypersurfaces are the simplest algebraic varieties—their cohomology looks exactly like that of $\mathbb{P}^k$ except in the middle dimension. It is a remarkable experience to actually see this cohomological behavior reflected in the Weil formulas for N_m, appearing almost out of nowhere. The second reason we include this material is to accentuate the fact that Weil's method is very *number-theoretic*. It is precisely this mysterious connection between number theory on the one hand, and algebraic topology on the other, which makes the Weil conjectures so wonderful and tantalizing.

B.1. Multiplicative characters. Before jumping into the calculation we need a simple tool. Let $F = \mathbb{F}_\nu$ be a finite field. Recall that the multiplicative group F^* is cyclic of order $\nu - 1$.

Fix a positive integer $d > 1$. The dth roots of unity in F constitute the kernel of the dth power map $F^* \to F^*$. Up to isomorphism this is $\mathbb{Z}/(\nu-1) \xrightarrow{d} \mathbb{Z}/(\nu-1)$, which has the same kernel as multiplication by e, where $e = (d, \nu - 1)$. Since $e|(\nu-1)$, this kernel evidently has e elements. Our conclusion is that the dth roots of unity in F coincide with the eth roots of unity, and that there are e of them. In particular, F contains all dth roots of unity precisely when $d|\nu - 1$.

If $u \in F$, let $\{u^{\frac{1}{d}}\}$ denote the number of dth roots of u in F. This number equals 1 if $u = 0$, it equals 0 if u is not a dth power, and if u *is* a dth power then it is equal to the number of dth roots of unity in F. Since the latter also equals the number of eth roots of unity in F, we have verified that

$$\{u^{\frac{1}{d}}\} = \{u^{\frac{1}{e}}\}$$

for any $u \in F$.

Recall that a **multiplicative character** is a group homomorphism $\chi\colon F^* \to \mathbb{C}^*$. Since F^* is finite, the image will necessarily lie inside the roots of unity in $\mathbb{C}$; and since F^* is cyclic, χ is completely determined by what it does to a generator g.

Let 1 denote the trivial character. For the moment we will mostly be considering characters $F^* \to \mu_d$, where μ_d denotes the group of dth roots of unity in $\mathbb{C}$. For such a character, $\chi(g)$ is both d-torsion and $(\nu-1)$-torsion, and hence it is in fact e-torsion (since e is a $\mathbb{Z}$-linear combination of d and $\nu-1$). So all characters $F^* \to \mu_d$ actually land inside of μ_e. Of course there are precisely e distinct characters $F^* \to \mu_e$.

By convention we set $\chi(0) = 0$ except when $\chi = 1$, in which case we set $\chi(0) = 1$. With these conventions one has that

$$\{u^{1/d}\} = \sum_{\chi\colon F^* \to \mu_d} \chi(u) \tag{B.2}$$

where the sum runs over all characters. To see why this works, first note that both sides remain the same upon replacing d by e. Let ζ be a primitive eth root of unity in $\mathbb{C}$. Write $u = g^k$, for some k, and then observe that the right-hand side is equal to $1 + \zeta^k + \zeta^{2k} + \cdots + \zeta^{(e-1)k}$. If $e|k$ then $\zeta^k = 1$ and this sum evidently equals e. If $e \nmid k$ then $\zeta^k \neq 1$, and this expression equals $\frac{(\zeta^k)^e - 1}{\zeta^k - 1}$; but this is zero, since $\zeta^e = 1$. The condition that $e|k$ is readily seen to be equivalent to u having an eth root in F, and so this completes the proof of (B.2).

Note that the above discussion is a bit easier in the case $d|\nu - 1$, only because we don't have to introduce e at all. This will play a role in the arguments below.

B.3. Counting points. Fix an integer $d > 1$, and fix a prime power q. Let X be the hypersurface over $\mathbb{F}_q$ defined by $x_0^d + \cdots + x_n^d = 0$, which is a projective variety in $\mathbb{P}^n$. Our goal is to compute $N_m(X)$, the number of points in X with values in $\mathbb{F}_{q^m}$. To make our calculation easier we will assume $d|q-1$, as this ensures that $\mathbb{F}_q$ (and all its extension fields) have a complete set of dth roots of unity.

Write $F = \mathbb{F}_{q^m}$, and let AN_m denote the number of *affine* solutions to the equation $x_0^d + \cdots + x_n^d = 0$ lying in F. So

$$AN_m = \sum_{u_0 + \cdots + u_d = 0} \{u_0^{\frac{1}{d}}\} \cdot \{u_1^{\frac{1}{d}}\} \cdots \{u_n^{\frac{1}{d}}\}$$

where the sum is taken over tuples $(u_0, \ldots, u_d) \in F^{d+1}$ and $\{u^{\frac{1}{d}}\}$ denotes the number of dth-roots of u in F. Using (B.2), we have

$$\begin{aligned} AN_m &= \sum_{u_0 + \cdots + u_n = 0} \left[\sum_{\chi_0, \ldots, \chi_n} \chi_0(u_0)\chi_1(u_1) \cdots \chi_n(u_n) \right] \\ &= \sum_{\chi_0, \ldots, \chi_n} \left[\sum_{u_0 + \cdots + u_n = 0} \chi_0(u_0)\chi_1(u_1) \cdots \chi_n(u_n) \right] \end{aligned} \tag{B.3}$$

where the characters χ_i are understood to take values in μ_d. The expression inside the brackets is a kind of *Jacobi sum*, which we will explore in (B.14) below. For now we just introduce the notation

$$J_0(\lambda_1, \ldots, \lambda_n) = \sum_{u_1 + \cdots + u_n = 0} \lambda_1(u_1)\lambda_2(u_2) \cdots \lambda_n(u_n).$$

Note that $J_0(1, 1, \ldots, 1) = (q^m)^{n-1}$.

LEMMA B.4. *Let $\chi_1, \ldots, \chi_n \colon \mathbb{F}_\nu \to \mathbb{C}$ be multiplicative characters.*

(a) If some of the χ_i's are trivial and some are nontrivial, then $J_0(\chi_1, \ldots, \chi_n) = 0$.

(b) If the product $\chi_1\chi_2 \cdots \chi_n$ is nontrivial, then $J_0(\chi_1, \ldots, \chi_n) = 0$.

PROOF. Both parts are based on the following observation. If χ is a nontrivial character on $\mathbb{F}_\nu$, then

$$\begin{aligned}\sum_{u\in\mathbb{F}_\nu}\chi(u) &= \chi(g)+\chi(g^2)+\cdots+\chi(g^{\nu-1})\\ &= [1+\chi(g)+\chi(g)^2+\cdots+\chi(g)^{\nu-1}]-1\\ &= \left[\frac{\chi(g)^\nu-1}{\chi(g)-1}\right]-1\\ &= \left[\frac{\chi(g^\nu)-1}{\chi(g)-1}\right]-1=\left[\frac{\chi(g)-1}{\chi(g)-1}\right]-1=0.\end{aligned}$$

We will prove (a) and (b) in the case $n=3$, and it will be clear how the general case follows. For (a), note that if χ_3 is nontrivial then

$$J_0(1,\chi_2,\chi_3)=\sum_{u_1+u_2+u_3=0}\chi_2(u_2)\chi_3(u_3)=\Big[\sum_{u_2\in F}\chi_2(u_2)\Big]\cdot\Big[\sum_{u_3\in F}\chi_3(u_3)\Big]=[??]\cdot 0=0.$$

For (b), let $\beta=\chi_1\chi_2\chi_3$. If $\beta\neq 1$ then at least one χ_i is nontrivial; assume it is χ_3. Then

$$\begin{aligned}J_0(\chi_1,\chi_2,\chi_3) &= \sum_{u_1+u_2+u_3=0}\chi_1(u_1)\chi_2(u_2)\chi_3(u_3)\\ &= \sum_{u_1+u_2+u_3=0,u_3\neq 0}\chi_1(u_1)\chi_2(u_2)\chi_3(u_3)\\ &= \sum_{u_1+u_2+u_3=0,u_3\neq 0}\chi_1(u_1)\chi_2(u_2)\cdot\frac{\beta(u_3)}{\chi_1(u_3)\chi_2(u_3)}\\ &= \sum_{u_1+u_2+u_3=0,u_3\neq 0}\chi_1\Big(\frac{u_1}{u_3}\Big)\chi_2\Big(\frac{u_2}{u_3}\Big)\beta(u_3)\\ &= \sum_{a+b+1=0,u_3\neq 0}\chi_1(a)\chi_2(b)\beta(u_3)\\ &= \Big[\sum_{a+b+1=0}\chi_1(a)\chi_2(b)\Big]\cdot\Big[\sum_{u\neq 0}\beta(u)\Big].\end{aligned}$$

If β is nontrivial then we know that $\sum_{u\neq 0}\beta(u)=\sum_u\beta(u)=0$, and hence $J_0(\chi_1,\chi_2,\chi_3)=0$. $\square$

At this point we have shown that many terms vanish in the sum (B.3). What we have left is

$$AN_m=(q^m)^n+\sum_{\chi_i\neq 1,\prod_i\chi_i=1}J_0(\chi_0,\ldots,\chi_n) \tag{B.5}$$

where the characters χ_i have the form $\mathbb{F}_{q^m}\to\mu_d$. To analyze the J_0 terms further, we will need the norm function $N=N_{\mathbb{F}_{q^m}/\mathbb{F}_q}\colon\mathbb{F}_{q^m}^*\to\mathbb{F}_q^*$ given by

$$N(x)=x\cdot x^q\cdot x^{q^2}\cdots x^{q^{m-1}}=x^{1+q+q^2+\cdots+q^{m-1}}.$$

This is a homomorphism of multiplicative groups, and it is actually surjective. To see this, note that the kernel of N is the set of roots of the polynomial $x^{\frac{q^m-1}{q-1}}-1$, and so the number of elements in the kernel is less than or equal to $\frac{q^m-1}{q-1}$. Since the domain has $q^m - 1$ elements, the image must therefore have at least $q-1$ elements. So the image must encompass all of $\mathbb{F}_q^*$.

If $\lambda\colon \mathbb{F}_q \to \mathbb{C}$ is a multiplicative character, then $\lambda \circ N$ is a multiplicative character for $\mathbb{F}_{q^m}$. Denote this character by $\lambda^{(m)}$. Our assumption that $d|q-1$ shows that every character $\mathbb{F}_{q^m} \to \mu_d$ has the form $\lambda^{(m)}$, for some $\lambda\colon \mathbb{F}_q \to \mu_d$ (this uses the fact that N is surjective). So we may rewrite (B.5) as

$$AN_m = (q^m)^n + \sum_{\lambda_i \neq 1, \prod_i \lambda_i = 1} J_0(\lambda_0^{(m)}, \dots, \lambda_n^{(m)}) \tag{B.6}$$

where the sum ranges over all characters $\lambda_i\colon \mathbb{F}_q \to \mu_d$.

We wish to compare $J_0(\chi_1, \dots, \chi_n)$ to $J_0(\chi_1^{(m)}, \dots, \chi_n^{(m)})$, and to do this it turns out to be convenient to introduce an auxilliary definition. Given characters $\lambda_i\colon \mathbb{F}_\nu \to \mathbb{C}$, define

$$j(\lambda_1, \dots, \lambda_n) = (-1)^n \frac{J_0(\lambda_1, \dots, \lambda_n)}{\nu - 1}.$$

The following result concerns this j function. It has a slightly involved proof, which depends on some very clever manipulations with Gauss sums (introduced below). For the moment we will defer the proof, and instead focus on how the result allows us to complete our calculation of the numbers N_m.

Theorem B.7. *Let $\chi_1, \chi_2, \dots, \chi_k\colon \mathbb{F}_q \to \mathbb{C}$ be nontrivial multiplicative characters. Then*
(a) $j(\chi_1, \dots, \chi_k)$ is an algebraic integer of norm $q^{\frac{k-2}{2}}$.
(b) $j(\overline{\chi_1}, \dots, \overline{\chi_k}) = \overline{j(\chi_1, \dots, \chi_k)}$.
(c) $j(\chi_1^{(m)}, \dots, \chi_k^{(m)}) = [j(\chi_1, \dots, \chi_k)]^m$. That is,

$$(-1)^k \cdot \frac{J_0(\chi_1^{(m)}, \dots, \chi_k^{(m)})}{q^m - 1} = \left[(-1)^k \frac{J_0(\chi_1, \dots, \chi_k)}{q-1}\right]^m.$$

Proof. See Section B.14 below. □

Returning now to equation (B.6), Theorem B.7(c) lets us make the substitution $J_0(\lambda_0^{(m)}, \dots, \lambda_n^{(m)}) = (-1)^{n+1}(q^m - 1) j(\chi_0, \dots, \chi_n)^m$. Finally, recall that we are really interested in counting the number of *projective* solutions of our equation rather than the number of affine solutions. Using $N_m = (AN_m - 1)/(q^m - 1)$ we get

$$N_m = [(q^m)^{n-1} + \cdots + q + 1] + (-1)^{n+1} \sum_{\lambda_i \neq 1, \prod_i \lambda_i = 1} j(\lambda_0, \dots, \lambda_n)^m \tag{B.8}$$

where the sum runs over characters $\lambda_i\colon \mathbb{F}_q \to \mu_d$. Note that in the above equation we have finally removed all references to $\mathbb{F}_{q^m}$.

B.9. A special case: elliptic curves in $\mathbb{P}^2$. For the moment we will now restrict to the case $n = 2$ and $d = 3$. That is, X is the subvariety of $\mathbb{P}^2$ defined by $x^3 + y^3 + z^3 = 0$. We are working over a field $\mathbb{F}_q$ where $3|q-1$.

Notice that there are exactly three characters $\mathbb{F}_q^* \to \mu_3$, as a multiplicative generator can be sent to any of the three cube roots of unity. Let g denote a chosen generator for $\mathbb{F}_q^*$, let $\zeta = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, and let χ denote the character sending g to ζ. Let $\bar{\chi}$ denote the character sending g to $\bar{\zeta}$.

There are only two ways to give three non-trivial characters χ_1, χ_2, χ_3 with $\prod \chi_i = 1$: one can have $\chi_1 = \chi_2 = \chi_3 = \chi$ or $\chi_1 = \chi_2 = \chi_3 = \bar{\chi}$. So (B.8) reduces to

$$N_m = q^m + 1 - [A^m + B^m]$$

where $A = j(\chi, \chi, \chi)$ and $B = j(\bar{\chi}, \bar{\chi}, \bar{\chi})$. Note that this is the form of N_m expected by the Weil conjectures, and that we have $|A| = |B| = \sqrt{q}$ by Theorem B.7(a)—thereby confirming the Riemann hypothesis in this case. Also, since $\bar{A} = B$ by Theorem B.7(b), we have $A = q/\bar{A} = q/B$, and this verifies Poincaré Duality.

Now we will choose specific values for q and compute the numbers A and B explicitly. Take $q = 7$ to start with, and let $g = 3$ be our chosen generator for $\mathbb{F}_q^*$. We must compute $J_0(\chi, \chi, \chi) = \sum_{u_1+u_2+u_3=0} \chi(u_1)\chi(u_2)\chi(u_3)$, where $u_1, u_2, u_3 \in \mathbb{F}_7$. If any $u_i = 0$ then $\chi(u_i) = 0$ and we can neglect that term. So there are really 30 terms in the sum: six non-zero choices for u_1, and then u_2 can be chosen to be anything in $\mathbb{F}_7 - \{0, -u_1\}$. Going through these 30 terms by brute force, we find that 12 of them have $\chi(u_1)\chi(u_2)\chi(u_3) = 1$ and 18 of them have $\chi(u_1)\chi(u_2)\chi(u_3) = \bar{\zeta}$. So

$$J_0(\chi, \chi, \chi) = 12 + 18\bar{\zeta}$$

and

$$j(\chi, \chi, \chi) = -\frac{12 + 18\bar{\zeta}}{7 - 1} = -\frac{1 - 3\sqrt{3}i}{2}.$$

Recall $A = j(\chi, \chi, \chi)$ and $B = j(\bar{\chi}, \bar{\chi}, \bar{\chi}) = \bar{A}$. That is,

$$A = \frac{-1 + 3\sqrt{3}i}{2} \qquad \text{and} \qquad B = \frac{-1 - 3\sqrt{3}i}{2}.$$

The same computations can be made with other values for q. A computer is useful for the brute force enumerations at the end. One finds the following, for example:

q	$J_0(\chi, \chi, \chi)$	A	$Z(X,t)$
7	$12 + 18\bar{\zeta}$	$\frac{-1+3\sqrt{3}i}{2}$	$\frac{1-t+7t^2}{(1-t)(1-7t)}$
13	$24 + 36\zeta + 72\bar{\zeta}$	$\frac{5+3\sqrt{3}i}{2}$	$\frac{1+5t+13t^2}{(1-t)(1-13t)}$
19	$144 + 54\zeta + 108\bar{\zeta}$	$\frac{7-3\sqrt{3}i}{2}$	$\frac{1+7t+19t^2}{(1-t)(1-11t)}$
31	$330 + 360\zeta + 180\bar{\zeta}$	$2 + 3\sqrt{3}i$	$\frac{1+4t+31t^2}{(1-t)(1-31t)}$
37	$288 + 540\zeta + 432\bar{\zeta}$	$\frac{-11-3\sqrt{3}i}{2}$	$\frac{1-11t+37t^2}{(1-t)(1-37t)}$
43	$462 + 756\zeta + 504\bar{\zeta}$	$-4 + 3\sqrt{3}i$	$\frac{1-8t+43t^2}{(1-t)(1-43t)}$

B.10. The cohomology of complex hypersurfaces. Our next task is to generalize the above example to all diagonal hypersurfaces, which means explaining how (B.8) meets the criteria of the Weil conjectures. Since the conjectures relate the number of points of varieties over finite fields to topological invariants of associated complex varieties, we will need to know a little about the topology of these hypersurfaces. If X is a degree d hypersurface in $\mathbb{C}P^n$, its cohomology groups are completely determined by d. This will be explained in more detail in Chapter 2, so for now we will be content to just state the facts.

Except for the middle dimension $n-1$, the cohomology groups of X are equal to $\mathbb{Z}$ in every even dimension between 0 and $2(n-1)$, and are equal to 0 in every odd dimension. In the middle dimension, $H^{n-1}(X) \cong \mathbb{Z}^{R_d}$ where

$$R_d = \begin{cases} \frac{(d-1)^{n+1}-(d-1)}{d} & \text{if } n \text{ is even,} \\ \frac{(d-1)^{n+1}+2d-1}{d} & \text{if } n \text{ is odd.} \end{cases}$$

The number R_d can also be written as

$$R_d = \begin{cases} R'_d & \text{if } n \text{ is odd,} \\ R'_d + 1 & \text{if } n \text{ is even.} \end{cases}$$

where

$$R'_d = \frac{1}{d}\big[(d-1)^{n+1} + (-1)^{n+1}(d-1)\big].$$

This can be interpreted as saying that $H^*(X)$ consists of a $\mathbb{Z}$ in every even dimension between 0 and $2n-2$, with an *extra* R'_d copies of $\mathbb{Z}$ in the middle dimension $n-1$.

Finally, we remark that the numbers R'_d satisfy the recurrence relation $R'_d + R'_{d-1} = (d-1)^n$, and so an easy induction yields

$$R'_d = (d-1)^n - (d-1)^{n-1} + \cdots + (-1)^{n-1}(d-1).$$

These different ways of looking at the number R'_d will be important below.

B.11. The general case of diagonal hypersurfaces. Now we return our analysis. Recall we have fixed q, and $X \hookrightarrow \mathbb{P}^n$ is the hypersurface defined by the equation $x_0^d + \cdots + x_n^d = 0$. Under the assumption $d|q-1$ we have shown that

(B.11)
$$N_m = [(q^{n-1})^m + (q^{n-2})^m + \cdots + q^m + 1] + (-1)^{n+1} \sum_{\lambda_i \neq 1, \prod_i \lambda_i = 1} j(\lambda_0, \ldots, \lambda_n)^m.$$

where the summation ranges over all characters $\lambda_i \colon \mathbb{F}_q \to \mu_d$.

Let C_n denote the number of $(n+1)$-tuples of characters $(\lambda_0, \ldots, \lambda_n)$ such that $\lambda_i \colon \mathbb{F}_q \to \mu_d$, each $\lambda_i \neq 1$, and $\lambda_0\lambda_1 \cdots \lambda_n = 1$; in other words, C_n is the number of terms in the summation part of (B.11). Also, let D_n denote the number of $(n+1)$-tuples satisfying $\lambda_i \neq 1$ and $\prod_i \lambda_i \neq 1$. Clearly $C_n + D_n = (d-1)^{n+1}$. One also has $C_n = D_{n-1}$, using the correspondence which assigns an n-tuple $(\lambda_0, \ldots, \lambda_{n-1})$ with $\prod \lambda_i \neq 1$ to the $(n+1)$-tuple $(\lambda_0, \ldots, \lambda_n, (\prod_i \lambda_i)^{-1})$. Hence we have the recurrence relation $C_n = (d-1)^n - C_{n-1}$, so that

$$C_n = (d-1)^n - (d-1)^{n-1} + (d-1)^{n-2} - \cdots + (-1)^n(d-1).$$

This is precisely the number R'_d from the previous section.

Recall from Section B.10 that the cohomology groups of X consist of a $\mathbb{Z}$ in every even dimension from 0 through $2(n-1)$, together with an extra R'_d copies of $\mathbb{Z}$ in the middle dimension $n-1$. Comparing this to equation (B.11), we see that the terms $1+q^m+\cdots+(q^{n-1})^m$ correspond to the former $\mathbb{Z}$'s, whereas the $C_n = R'_d$ terms inside the summation correspond to the 'extra' $\mathbb{Z}$'s in the middle dimension. By Theorem B.7(a), the norms of these terms inside the summation are precisely $[q^{(n-1)/2}]^m$, in agreement with the Riemann Hypothesis. Also by Theorem B.7(a), the numbers $j(\lambda_0, \ldots, \lambda_n)$ are algebraic integers. Finally, Poincaré Duality asks that the two sequences of numbers $\{j(\lambda_0, \ldots, \lambda_n)\}$ and $\{q^{n-1}/j(\lambda_0, \ldots, \lambda_n)\}$ be the same up to permutation. We see this by noting that $j(\lambda_0,\dots,\lambda_n)$ is equal to $q^{n-1}/j(\overline{\lambda_0},\dots,\overline{\lambda_n})$, by Theorem B.7(a,b).

Finally, consider the zeta function $Z(X,t)$ for our hypersurface. The formula (B.11) shows that this function has the form

$$Z(X,t) = \frac{P(t)^{(-1)^n}}{(1-t)(1-qt)(1-q^2t)\cdots(1-q^{n-1}t)} \tag{B.12}$$

where

$$P(t) = \prod_{\lambda_i\neq 1, \prod \lambda_i = 1} \Big(1 - j(\lambda_0,\dots,\lambda_n)t\Big).$$

The original definition of $Z(X,t)$ shows that its coefficients lie in $\mathbb{Q}$. Solving (B.12) for $P(t)$ then shows that the coefficients of $P(t)$ are also in $\mathbb{Q}$. However, the above product expansion tells us that the coefficients of $P(t)$ are also algebraic integers. The only rational numbers which are algebraic integers are the actual integers, so therefore $P(t) \in \mathbb{Z}[t]$. We have now verified all of the Weil conjectures for hypersurfaces defined by equations $x_0^d+\cdots+x_n^d = 0$ over $\mathbb{F}_q$, assuming that $d|q-1$.

EXERCISE B.13. Let X be the projective variety over $\mathbb{F}_q$ defined by the equation $a_0x_0^d + \cdots + a_nx_n^d = 0$, where $a_0,\dots,a_n \in \mathbb{F}_q^*$. Show that the number of affine solutions to this equation over $\mathbb{F}_{q^m}$ is

$$AN_m = \sum_{\chi_0,\dots,\chi_n\colon \mathbb{F}_{q^m}\to\mu_d} \chi_0(a_0^{-1})\cdots\chi_n(a_n^{-1})\cdot J_0(\chi_0,\dots,\chi_n).$$

Building off of the case $a_0 = a_1 = \cdots = a_n = 1$, show that the Weil conjectures hold for all diagonal hypersurfaces provided $d|q-1$.

B.14. Gauss sums and Jacobi sums. In this final section we turn to the proof of Theorem B.7, which was the key step in our calculation of N_m for diagonal hypersurfaces. The techniques of this proof will not be needed elsewhere in this document, and so this material is a bit of a digression. In order to keep it from being too *much* of a digression, we will not actually give the whole proof—the key step is the Hasse-Davenport relation from number theory, and for this we will just refer to an appropriate source. Still, it seems worthwhile to introduce Gauss and Jacobi sums, which are important tools in number theory and which serve to give some context to Weil's results. Our treatment has been heavily influenced by the one in [**IR**].

Consider a finite field $F = \mathbb{F}_q$, where $q = p^e$. Given $s \in F$ and characters $\chi_1,\dots,\chi_n\colon F\to\mathbb{C}$, define the **Jacobi sum** as

$$J_s(\chi_1,\dots,\chi_n) = \sum_{u_1+\cdots+u_n=s} \chi_1(u_1)\cdots\chi_n(u_n).$$

LEMMA B.15. *If $s\neq 0$ then $J_s(\chi_1,\dots,\chi_n) = (\chi_1\cdots\chi_n)(s)\cdot J_1(\chi_1,\dots,\chi_n)$.*

PROOF. One uses the change-of-variable $a_i = u_i/s$ to see that

$$\begin{aligned}\sum_{u_1+\cdots+u_n=s} \chi_1(u_1)\cdots\chi_n(u_n) &= \sum_{a_1+\cdots+a_n=1} \chi_1(s\cdot a_1)\cdots\chi_n(s\cdot a_n)\\ &= \chi_1(s)\cdots\chi_n(s)\cdot J_1(\chi_1,\dots,\chi_n).\end{aligned}$$

□

Next we introduce Gauss sums. An **additive character** is a homomorphism from the additive group of F to the multiplicative group $\mathbb{C}^*$. If we choose in advance a primitive pth root of unity ζ, then we can construct a canonical additive character $\psi\colon \mathbb{F}_q \to \mathbb{C}$ by

$$\psi(x) = \zeta^{\mathrm{tr}(x)}$$

where $\mathrm{tr} = \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}\colon \mathbb{F}_q \to \mathbb{F}_p$ is the usual trace function, given by

$$\mathrm{tr}(x) = x + x^p + x^{p^2} + \cdots + x^{p^{e-1}}.$$

The additive character ψ is 'canonical' in the sense that the same formula works for all extension fields of $\mathbb{F}_p$.

If $\chi\colon F \to \mathbb{C}$ is a multiplicative character, define the **Gauss sum** of χ to be

$$g(\chi) = \sum_{x\in F} \chi(x)\psi(x).$$

Note that all the terms $\chi(x)$ and $\psi(x)$ are algebraic integers (being roots of unity), and so every Gauss sum is an algebraic integer.

The following two theorems contain most of what we will need about Gauss sums:

Theorem B.16. *Let χ be any multiplicative character for $F = \mathbb{F}_q$. Then*
(a) $g(\chi^{-1}) = \chi(-1)\overline{g(\chi)}$.
(b) If χ is nontrivial then $g(\chi)\overline{g(\chi)} = q$.

Proof. For (a), note that $\chi^{-1} = \overline{\chi}$, and that for any $t \in F$ one has $\overline{\psi(t)} = \psi(t)^{-1} = \psi(-t)$. Then

$$\begin{aligned}\overline{g(\chi)} = \sum_{t\in F} \overline{\chi(t)}\cdot\overline{\psi(t)} = \sum_t \overline{\chi(t)}\cdot\psi(-t) &= \sum_s \overline{\chi(-s)}\psi(s) \\ &= \sum_s \chi^{-1}(-1)\cdot\chi^{-1}(s)\psi(s) \\ &= \chi^{-1}(-1)\cdot g(\chi^{-1}).\end{aligned}$$

For (b) one first introduces the auxilliary sums $g_\alpha(\chi) = \sum_t \chi(t)\psi(\alpha t)$. One readily checks that $g_0(\chi) = 0$ and that $g_\alpha(\chi) = \chi(\alpha^{-1})g(\chi)$ for $\alpha \neq 0$. Now consider the sum $A = \sum_\alpha g_\alpha(\chi)\overline{g_\alpha(\chi)}$. One the one hand we have

$$A = \sum_\alpha \chi(\alpha^{-1})g(\chi)\cdot\overline{\chi(\alpha^{-1})}\cdot\overline{g(\chi)} = (q-1)g(\chi)\overline{g(\chi)}.$$

Looking at it another way, we have that

$$A = \sum_\alpha \Big[\sum_t \chi(t)\psi(\alpha t)\Big]\cdot\Big[\sum_s \overline{\chi(s)\psi(\alpha s)}\Big] = \sum_{t,s} \chi(t)\overline{\chi(s)}\sum_\alpha \psi(\alpha(t-s)).$$

Consider the term $\sum_\alpha \psi(\alpha(t-s))$. If $t-s=0$ then this sum is $|F| = q$; if $t-s \neq 0$ then by a change of variable it is just $\sum_\beta \psi(\beta)$, which is 0 by Lemma B.17 below. Using these observations, we now have

$$A = \sum_t \chi(t)\overline{\chi(t)}\cdot q = \Big(\sum_{t\neq 0} 1\Big)\cdot q = (q-1)q.$$

Comparing our two formulas for A, we find that $g(\chi)\overline{g(\chi)} = q$. □

The following lemma was used in the above proof:

LEMMA B.17. *$\sum_{t\in F}\psi(t) = 0$.*

PROOF. Let $B = \sum_t \psi(t)$. For any $x \in F$ we have that

$$\psi(x)\cdot B = \sum_{t\in F}\psi(x+t) = \sum_{u\in F}\psi(u) = B.$$

But it is easy to see that there exists an x such that $\psi(x) \neq 1$, and therefore B must be zero. □

THEOREM B.18 (Hasse-Davenport relation). *For any character $\chi\colon \mathbb{F}_q \to \mathbb{C}$, one has*

$$g(\chi^{(m)}) = (-1)^{m+1}\cdot g(\chi).$$

PROOF. See [**IR**, Chapter 11.4]. The proof is elementary, but somewhat too long to include here. □

Now we turn to the connection between Gauss and Jacobi sums.

THEOREM B.19. *Let $\chi_1,\ldots,\chi_n$ be multiplicative characters $\mathbb{F}_q \to \mathbb{C}$.*

(a) $g(\chi_1)\cdots g(\chi_n) = J_0(\chi_1,\ldots,\chi_n) + J_1(\chi_1,\ldots,\chi_n)[g(\chi_1\cdots\chi_n) - (\chi_1\cdots\chi_n)(0)]$.

(b) If $\prod_i \chi_i \neq 1$ then $J_1(\chi_1,\ldots,\chi_n) = \dfrac{g(\chi_1)\cdots g(\chi_n)}{g(\chi_1\cdots\chi_n)}$.

(c) If all $\chi_i \neq 1$ and $\prod_i \chi_i = 1$, then

$$J_0(\chi_1,\ldots,\chi_n) = (q-1)\cdot\Big[\frac{g(\chi_1)\cdots g(\chi_n)}{q}\Big].$$

PROOF. We start with (a).

$$\begin{aligned}
g(\chi_1)\cdots g(\chi_n) &= \Big[\sum_{u_1}\chi_1(u_1)\psi(u_1)\Big]\cdots\Big[\sum_{u_n}\chi_n(u_n)\psi(u_n)\Big]\\
&= \sum_s \sum_{u_1+\cdots+u_n=s}\chi_1(u_1)\cdots\chi_n(u_n)\psi(s)\\
&= \sum_s J_s(\chi_1,\ldots,\chi_n)\psi(s)\\
&= J_0(\chi_1,\ldots,\chi_n) + \sum_{s\neq 0} J_1(\chi_1,\ldots,\chi_n)\cdot(\chi_1\cdots\chi_n)(s)\cdot\psi(s)\\
&= J_0(\chi_1,\ldots,\chi_n) + J_1(\chi_1,\ldots,\chi_n)[g(\chi_1\cdots\chi_n) - (\chi_1\cdots\chi_n)(0)].
\end{aligned}$$

To prove part (b), first note that if $\chi_1\cdots\chi_n \neq 1$ then $(\chi_1\cdots\chi_n)(0) = 0$. Second, recall that we have already proven in Lemma B.4(b) that $J_0(\chi_1,\ldots,\chi_n) = 0$ if $\prod_i \chi_i \neq 1$. So (b) follows at once from (a).

To prove (c) we argue as follows. First,

$$\begin{aligned}
J_0(\chi_1,\dots,\chi_n) &= \sum_{u_1+\cdots+u_n=0} \chi_1(u_1)\cdots\chi_n(u_n) \\
&= \sum_{u_n\neq 0}\ \sum_{u_1+\cdots+u_n=0} \chi_1(u_1)\cdots\chi_n(u_n) \qquad (\text{since } \chi_n\neq 1) \\
&= \sum_{s\neq 0}\Big[\sum_{u_1+\cdots+u_{n-1}=-s} \chi_1(u_1)\cdots\chi_{n-1}(u_{n-1})\Big]\cdot\chi_n(s) \\
&= \sum_{s\neq 0} J_{-s}(\chi_1,\dots,\chi_{n-1})\cdot\chi_n(s) \\
&= \sum_{s\neq 0} J_1(\chi_1,\dots,\chi_{n-1})(\chi_1\cdots\chi_{n-1})(-s)\cdot\chi_n(-1)\chi_n(-s) \\
&= J_1(\chi_1,\dots,\chi_{n-1})\sum_{s\neq 0}\chi_n(-1) \qquad (\text{since } \chi_1\cdots\chi_n=1) \\
&= \chi_n(-1)\cdot(q-1)J_1(\chi_1,\dots,\chi_{n-1}).
\end{aligned}$$

But $\chi_1\cdots\chi_{n-1}=\chi_n^{-1}\neq 1$, and so by (b) we have that

$$\begin{aligned}
J_1(\chi_1,\dots,\chi_{n-1}) = \frac{g(\chi_1)\cdots g(\chi_{n-1})}{g(\chi_n^{-1})} = \frac{g(\chi_1)\cdots g(\chi_{n-1})}{\chi_n(-1)\cdot\overline{g(\chi_n)}} &= \frac{g(\chi_1)\cdots g(\chi_n)}{\chi_n(-1)\cdot\overline{g(\chi_n)}g(\chi_n)} \\
&= \frac{g(\chi_1)\cdots g(\chi_n)}{\chi_n(-1)\cdot q}.
\end{aligned}$$

We have used Theorem B.16 in the second and fourth equalities. Putting everything together we now have $J_0(\chi_1,\dots,\chi_n)=(q-1)\cdot g(\chi_1)\cdots g(\chi_n)/q$. □

Finally, we close with the

Proof of Theorem B.7. We have that

$$j(\chi_1,\dots,\chi_n) = (-1)^n\frac{J_0(\chi_1,\dots,\chi_n)}{q-1} = (-1)^n\frac{g(\chi_1)\cdots g(\chi_n)}{q}$$

using the preceding theorem. So

$$|j(\chi_1,\dots,\chi_n)| = \frac{1}{q}\cdot|g(\chi_1)|\cdots|g(\chi_n)| = \frac{1}{q}\cdot q^{n/2} = q^{(n-2)/2}.$$

Also, during the proof of Theorem B.19 we showed that $j(\chi_1,\dots,\chi_n)$ is equal to $(-1)^n\chi_n(-1)J_1(\chi_1,\dots,\chi_n)$, and the latter is manifestly an algebraic integer. We also have

$$\begin{aligned}
j(\overline{\chi_1},\dots,\overline{\chi_n}) = (-1)^n\frac{g(\overline{\chi_1})\cdots g(\overline{\chi_n})}{q} &= (-1)^n\chi_1(-1)\cdots\chi_n(-1)\cdot\frac{\overline{g(\chi_1)\cdots g(\chi_n)}}{q} \\
&= \overline{j(\chi_1,\dots,\chi_n)}
\end{aligned}$$

where in the last equality we have used $\chi_1\cdots\chi_n=1$. Finally, Theorem B.7(c) is a direct consequence of Theorem B.19(c) and the Hasse-Davenport relation. □

CHAPTER 2

Topological interlude: the cohomology of algebraic varieties

This chapter represents a brief detour. Our goal is to review some basic facts about the topology of complex algebraic varieties. This material will be applied in the next chapter, when we return to the Weil conjectures and the search for a cohomology theory for varieties in characteristic p.

Given a smooth, compact algebraic variety over $\mathbb{C}$, what do its singular cohomology groups look like? Of course they must satisfy Poincaré Duality, but it turns out one can say much more. The pioneering work on this topic was done by Lefschetz [**L**], but that book is hard to read from a modern perspective—and some of the proofs may be incomplete. Lefschetz's theorems have been reproven over the years, and expanded on, using Morse and Hodge theory. Hodge's techniques [**Ho**] have been particularly important.

The first four sections of this chapter review Lefschetz and Hodge theory, just giving the basic facts without proof. In the final chapter we review the main ideas of *correspondences* (also due to Lefschetz, actually). These apply not just to varieties but to all compact manifolds.

1. Lefschetz theory

Let X be a complex projective algebraic variety. By a **hyperplane section** of X one means any variety of the form $X \cap H$ where H is a hyperplane in $\mathbb{C}P^n$ and $X \hookrightarrow \mathbb{C}P^n$ is some embedding. Lefschetz was interested in studying the topology of X via the topology of its hyperplane sections.

REMARK 1.1. Suppose $X \hookrightarrow \mathbb{C}P^n$ and Z is a hypersurface in $\mathbb{C}P^n$. Then $X \cap Z$ is a hyperplane section of X. To see this, recall that Z can be defined by the vanishing of a single homogeneous polynomial $f \in \mathbb{C}[X_0, \ldots, X_n]$. Let d be the degree of f. Let $\{M_0, \ldots, M_N\}$ be a complete list of the degree d monomials in the X_i's, where $N = \binom{n+d}{n} - 1$. Write $f = \sum_j a_j M_j$. Finally, recall the Veronese embedding $\mathbb{C}P^n \hookrightarrow \mathbb{C}P^N$ given by sending a point $x = [x_0 : x_1 : \cdots : x_N]$ to the sequence of monomials $[M_0(x) : M_1(x) : \ldots : M_N(x)]$.

We now have $X \hookrightarrow \mathbb{C}P^n \hookrightarrow \mathbb{C}P^N$. Let H be the hypersurface in $\mathbb{C}P^N$ defined by $a_0 Y_0 + a_1 Y_1 + \cdots + a_N Y_N = 0$, where the Y_i's are the evident homogeneous coordinates on $\mathbb{C}P^N$. One checks that $X \cap Z$ is homeomorphic to $X \cap H$.

THEOREM 1.2 (Weak Lefschetz). *Let W be a smooth, connected, projective, complex algebraic variety of dimension $n+1$, and let $X \hookrightarrow W$ be a smooth hyperplane section (so $\dim X = n$). Then the following statements hold:*

(a) $H_*(X) \to H_*(W)$ is an isomorphism for $i < n$ and a surjection for $i = n$.
(b) $H^*(W) \to H^*(X)$ is an isomorphism for $i < n$ and an injection for $i = n$.

(c) $\pi_*(X) \to \pi_*(W)$ *is an isomorphism for* $i < n$ *and a surjection for* $i = n$.

To remember the above result, note that each part says that $X \hookrightarrow W$ induces isomorphisms up through (but not including) the middle dimension of X. Parts (a) and (b) are equivalent, and are consequences of (c). The best proofs of this theorem seem to be via Morse theory. Proofs of (a) and (b) can be found in [**AF**] and [**Mr1**]. For (c) one must look to Bott [**B**].

EXAMPLE 1.3 (The cohomology of hypersurfaces). Let $X \hookrightarrow \mathbb{P}^k$ be a smooth hypersurface. In particular, X is a hyperplane section of $\mathbb{P}^k$. The Weak Lefschetz Theorem says that $H^*(\mathbb{P}^k) \to H^*(X)$ is an isomorphism for $* < k-1$. By Poincaré Duality for X, this completely determines $H^*(X)$ except for $* = k-1$.

Poincaré Duality also gives that the torsion subgroup of $H^i(X)$ is isomorphic to the torsion subgroup of $H^{2k-3-i}(X)$, and in particular that the torsion subgroups of $H^{k-1}(X)$ and $H^{k-2}(X)$ are isomorphic. But as $H^{k-2}(X) \cong H^{k-2}(\mathbb{P}^k)$, it has no torsion. So $H^{k-1}(X)$ is free abelian.

We can compute the rank of $H^{k-1}(X)$ if we know the Euler characteristic $\chi(X)$, since we know all the other cohomology groups. To be precise, one has

$$\operatorname{rank} H^{k-1}(X) = \begin{cases} k - \chi(X) & \text{if } k-1 \text{ is odd} \\ \chi(X) - k + 1 & \text{if } k-1 \text{ is even.} \end{cases}$$

The Euler characteristic of a hypersurface may be computed by the Hirzebruch-Riemann-Roch Theorem, and it turns out to only depend on the degree d of the hypersurface. Using [**H**, Thm. 22.1.1] in conjunction with [**H**, Thm. 15.8.1] one finds that $\chi(X)$ for a degree d hypersurface in $\mathbb{P}^k$ is the coefficient of z^k in the formal power series

$$\frac{1}{(1-z)^2} \cdot \frac{d \cdot z}{1 + (d-1)z}.$$

Some calculating shows this number to be

$$\chi(X) = \frac{(1-d)^{k+1} - 1}{d} + k + 1.$$

Putting everything together, one has

$$\operatorname{rank} H^{k-1}(X) = \begin{cases} \frac{(d-1)^{k+1} - d + 1}{d} & \text{if } \dim X \text{ is odd} \\ \frac{(d-1)^{k+1} + 2d - 1}{d} & \text{if } \dim X \text{ is even.} \end{cases}$$

As an example of the above, let's consider the surface $x^3 + y^3 + z^3 + u^3 + w^3 = 0$ in $\mathbb{C}P^4$. This is 3-dimensional and it has degree 3. We have completely determined its cohomology groups:

i	0	1	2	3	4	5	6
$H^i(X)$	$\mathbb{Z}$	0	$\mathbb{Z}$	$\mathbb{Z}^{10}$	$\mathbb{Z}$	0	$\mathbb{Z}$

EXAMPLE 1.4 (The cohomology of complete intersections). Recall that a complete intersection is a subvariety of $\mathbb{C}P^n$ defined by the vanishing of homogeneous polynomials $f_1, \ldots, f_k \in \mathbb{C}[x_0, \ldots, x_n]$ such that the f_i's are a regular sequence. Applying the Weak Lefschetz Theorem inductively, one finds that if X is a complete intersection then $H^i(\mathbb{C}P^n) \to H^i(X)$ is an isomorphism for $i < \dim X$. If X is smooth, this determines $H^i(X)$ for $i > \dim X$ by Poincaré Duality. The only unknown cohomology group is the middle one.

Just as before, Poincaré duality now shows that this middle cohomology group is torsion free. So its rank is completely determined by $\chi(X)$, which can be computed using the Hirzebruch-Riemann-Roch theorem. If $d_i = \deg f_i$, then the Euler characteristic is the coefficient of z^n in the series

$$\frac{1}{(1-z)^2} \cdot \prod_i \left[\frac{d_i z}{1+(d_i-1)z} \right]$$

(by [**H**, Thm. 22.1.1] and [**H**, Thm. 15.8.1] again). These coefficients are very computable in any specific case, but general formulas become unwieldy beyond this point.

As a specific example, suppose X is a complete intersection in $\mathbb{C}P^5$ defined by two forms, of degrees 2 and 3. One can use Mathematica to expand the above power series. The commands to do this are as follows:

```
In[1]:=   f[z_]:=(2*z/(1+z))*(3*z/1+2z)*1/(1-z)^2
In[2]:=   Series[f[z],{z,0,7}]
```

The "`In[1]:`" and "`In[2]:`" are Mathematica prompts, not to be entered by the user. Mathematica will return the power series $6z^2-6z^3+24z^4-36z^5+90z^6-162z^7$, and we are interested in the coefficient of z^5.

We know that X is a 3-dimensional complex projective variety whose cohomology agrees with $\mathbb{P}^3$ except in the middle dimension, where it is free abelian. The above power series calculation gives $\chi(X) = -36$, and this shows $H^*(X)$ is as follows:

i	0	1	2	3	4	5	6
$H^i(X)$	$\mathbb{Z}$	0	$\mathbb{Z}$	$\mathbb{Z}^{40}$	$\mathbb{Z}$	0	$\mathbb{Z}$

2. The Hard Lefschetz theorem

The complex structure on $\mathbb{C}P^n$ determines an orientation, which determines the Poincaré Duality isomorphism. We let $\xi \in H^2(\mathbb{C}P^n)$ denote the Poincaré dual to the fundamental class $[\mathbb{C}P^{n-1}] \in H_{2n-2}(\mathbb{C}P^n)$.

Let $X \hookrightarrow \mathbb{C}P^n$ be a smooth subvariety of dimension r. The image of ξ under the map $H^2(\mathbb{C}P^n) \to H^2(X)$ will also be denoted ξ, by abuse. We define the Lefschetz operator $L\colon H^i(X) \to H^{i+2}(X)$ by $L(x) = x \cdot \xi$. The class $\xi \in H^2(X)$ is often called a *hyperplane section* of X; under Poincaré Duality it corresponds to the fundamental class of $[X \cap \mathbb{C}P^{n-1}] \in H_{2(n-1)}(X)$ for a sufficiently general $\mathbb{C}P^{n-1}$ in $\mathbb{C}P^n$.

Theorem 2.1 (Hard Lefschetz). *Let $X \hookrightarrow \mathbb{C}P^n$ be a smooth subvariety of dimension r. Then the map $L^{r-i}\colon H^i(X;\mathbb{Q}) \to H^{2r-i}(X;\mathbb{Q})$ is an isomorphism, for every i in the range $0 \le i \le r$.*

Remark 2.2. The Hard Lefschetz theorem is not true with integral coefficients. As one example, consider the quadric $Q \hookrightarrow \mathbb{C}P^4$ defined by $x^2+y^2+z^2+w^2 = 0$. The cohomology ring is $H^*(Q) \cong \mathbb{Z}[x,y]/(x^2 = 2y, y^2)$ where x has degree 2 and y has degree 4. The map $L\colon H^2(Q) \to H^4(Q)$ sends a generator to twice a generator, and so is not an isomorphism.

One has the following simple corollary:

COROLLARY 2.3. *Let X be a smooth, compact, algebraic variety over $\mathbb{C}$. Then the even (resp. odd) Betti numbers of X are monotone increasing up through the middle dimension. That is, one has*

$$\beta_0 \leq \beta_2 \leq \beta_4 \leq \cdots \qquad \text{and} \qquad \beta_1 \leq \beta_3 \leq \beta_5 \leq \cdots$$

with the chain of inequalities stopping (or reversing itself, if you like) after passing $\beta_{\dim X}$.

It follows from the Hard Lefschetz theorem that for $i \leq r$ one can decompose the group $H^i(X;\mathbb{Q})$ into two pieces. One piece is the image of $L\colon H^{i-2}(X;\mathbb{Q}) \to H^i(X;\mathbb{Q})$ (which is an injection, by the above theorem) and represents the 'uninteresting' part of $H^i(X;\mathbb{Q})$. The other piece is called the *primitive* part of $H^i(X)$: one defines

$$PH^i(X;\mathbb{Q}) = \ker\Big[L^{r-i+1}\colon H^i(X;\mathbb{Q}) \to H^{2r-i+2}(X;\mathbb{Q})\Big].$$

It is an easy exercise to verify that one has a direct sum decomposition

$$H^i(X;\mathbb{Q}) = PH^i(X;\mathbb{Q}) \oplus \operatorname{im} L.$$

It follows that there is a decomposition

$$H^i(X;\mathbb{Q}) = PH^i(X;\mathbb{Q}) \oplus L[PH^{i-2}(X;\mathbb{Q})] \oplus L^2[PH^{i-4}(X;\mathbb{Q})] \oplus \cdots$$

This is called the *Lefschetz primitive decomposition* for $H^*(X;\mathbb{Q})$. Note that it depends on the embedding $X \hookrightarrow \mathbb{C}P^n$, as that is what determines the class ξ. For this reason the decomposition is not natural in X.

Because we will need it in the next chapter, we briefly mention the Lefschetz Λ-operator. This is a map $\Lambda\colon H^i(X;\mathbb{Q}) \to H^{i-2}(X;\mathbb{Q})$ (defined for all i) which in some ways plays the role of an inverse to L. Specifically, one has $\Lambda L = id$ on $H^i(X;\mathbb{Q})$ if $i <= \dim X - 2$, and $L\Lambda = id$ on $H^i(X;\mathbb{Q})$ if $i \geq \dim X + 2$. It is clear how to define Λ on each of the pieces of the primitive decomposition: for $a \in H^i(X;\mathbb{Q})$, define

$$\Lambda(a) = \begin{cases} 0 & \text{if } a \in PH^i(X;\mathbb{Q}), \\ L^{j-1}a & \text{if } a \neq 0 \text{ and } a = L^j b, \text{ for some } b \in PH^{i-2j}(X) \text{ and some } j \geq 1. \end{cases}$$

EXERCISE 2.4 (The Λ^c-operator). This exercise concerns a different way of viewing the Lefschetz primitive decomposition of $H^*(X;\mathbb{Q})$. Consider the Lie algebra $\mathfrak{sl}_2(\mathbb{Q})$ of trace zero 2×2 matrices. This is three-dimensional over $\mathbb{Q}$ with generators

$$e = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \text{and} \qquad h = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

satisfying the commutation relations

$$[e,f] = h, \qquad [e,h] = -2e, \qquad [f,h] = 2f.$$

For each integer n, define an irreducible $\mathfrak{sl}_2(\mathbb{Q})$-module $W(n)$ as follows: it has dimension $n+1$ and generators $w_n, w_{n-2}, w_{n-4}, \ldots, w_{-n}$, subject to the relations

$$f.w_i = w_{i+2}, \qquad h.w_i = iw_i, \quad \text{and} \quad e.w_{n-2i} = (i+1)(n-i)w_{n-2i+2}.$$

In writing these relations our convention is that $w_{n+2} = 0 = w_{-n-2}$. Note also that h acts diagonally on $W(n)$, with integral eigenvalues, and the i-eigenspace is the one-dimensional subspace spanned by w_i.

(a) Verify that $W(n)$ is an $\mathfrak{sl}_2(\mathbb{Q})$-module, and that it is irreducible.

(b) Let X be a smooth, projective algebraic variety of dimension d, and let $\xi \in H^2(X;\mathbb{Q})$ be the class of a hyperplane section. Define an operator $\Lambda^c\colon H^i(X;\mathbb{Q}) \to H^{i-2}(X;\mathbb{Q})$ by using the following formula and extending linearly:
$$\Lambda^c(a) = \begin{cases} 0 & \text{if } a \in PH^i(X;\mathbb{Q}), \\ j(d-i+j+1)L^{j-1}b & \text{if } a = L^j b \text{ for } b \in PH^{i-2j}(X;\mathbb{Q}),\ j \geq 1. \end{cases}$$
Verify that $H^*(X;\mathbb{Q})$ then becomes an $\mathfrak{sl}_2(\mathbb{Q})$-module via the formulas
$$f.a = La, \qquad e.a = \Lambda^c a, \qquad h.a = (i-d).a$$
for $a \in H^i(X;\mathbb{Q})$.

(c) Verify that the primitive decomposition of $H^*(X;\mathbb{Q})$ is the same as a decomposition into irreducible $\mathfrak{sl}_2(\mathbb{Q})$-modules. Specifically, if a_k denotes the dimension of $PH^k(X;\mathbb{Q})$ for $0 \leq k \leq d$, then as an $\mathfrak{sl}_2(\mathbb{Q})$-module $H^*(X;\mathbb{Q})$ is isomorphic to
$$\bigoplus_{0\leq k\leq d} W(d-k)^{\oplus a_k}.$$

3. The Hodge index theorem

Let X be a smooth, compact variety over $\mathbb{C}$, and let $r = \dim X$. The orientation on X determines an isomorphism $\eta\colon H^{2r}(X;\mathbb{Q}) \to \mathbb{Q}$. The cup product therefore induces a bilinear form on $H^r(X;\mathbb{Q})$ by setting $\langle a, b\rangle = \eta(a \cdot b)$. When r is even this form is symmetric, whereas when r is odd it is alternating. Poincaré Duality gives that the bilinear form is nondegenerate, since it says that the cup product $H^r(X;\mathbb{Q}) \otimes H^r(X;\mathbb{Q}) \to H^{2r}(X;\mathbb{Q})$ is a perfect pairing.

Theorem 3.1 (Hodge index theorem). *Suppose $r = 2j$. Then the symmetric bilinear form $\langle -, -\rangle$ on $PH^r(X;\mathbb{Q})$ is positive definite if j is even, and negative definite if j is odd.*

Example 3.2 (Cohomology ring of even-dimensional complete intersections). Let X be a complete intersection of complex dimension $2n$. By Weak Lefschetz we know that $H^{2i+1}(X) = 0$ for all i, and $H^{2i}(X) = \mathbb{Z}$ for $0 \leq i \leq 2n$ and $i \neq n$. We also know $H^{2n}(X) \cong \mathbb{Z}^k$ for some k, and we can determine the rank in terms of the degrees of the equations defining X.

Let $\xi \in H^2(X)$ be the class of a hyperplane section. The Hard Lefschetz theorem implies that ξ^j is a generator for $H^{2j}(X;\mathbb{Q})$ for all $0 \leq j \leq 2n$ except $j = n$.

Now, $PH^{2n}(X;\mathbb{Q})$ is by definition the kernel of multiplication by ξ. It is therefore a vector space of dimension $k-1$. We will have completely computed the ring structure on $H^*(X;\mathbb{Q})$ if we know it on $PH^{2n}(X;\mathbb{Q})$. While the Hodge Index Theorem does not completely calculate this for us, it does calculate the ring $H^*(X;\mathbb{R})$: it implies there exists a basis $b_1, \ldots, b_{k-1}$ for $PH^{2n}(X;\mathbb{R})$ such that $b_i \cdot b_j = (-1)^n \delta_{i,j}$.

For example, consider the quadric Q_{2n} given by $x_0^2 + x_1^2 + \cdots + x_{2n+1}^2 = 0$. In this case $H^{2n}(Q) \cong \mathbb{Z} \oplus \mathbb{Z}$. If $n = 2k$ then $H^*(Q;\mathbb{R}) \cong \mathbb{R}[x,y]/(x^{n+1}, xy, y^2 - x^n)$ where x has degree 2 and y has degree $2n$. If $n = 2k+1$ then $H^*(Q;\mathbb{R}) \cong \mathbb{R}[x,y]/(x^{n+1}, xy, y^2 + x^n)$ where the degrees of x and y are as before.

The following is an easy consequence of the Hodge Index Theorem, using Weak Lefschetz. It generalizes the index theorem to all cohomology groups, rather than just the middle-dimensional one.

COROLLARY 3.3. *Choose a hyperplane section of X and let L be the associated Lefschetz operator. For any j such that $2j \leq \dim X$, the symmetric bilinear form on $H^{2j}(X;\mathbb{Q})$ given by*

$$x, y \mapsto \eta(L^{r-2j}(x) \cdot y)$$

restricts to a positive definite form on $PH^{2j}(X;\mathbb{Q})$ if j is even, and a negative definite form if j is odd.

PROOF. The proof is by induction on $\dim X - 2j$. If this number equals zero, then the result is just Theorem 3.1. Otherwise, let Y be a hyperplane section of X, and let $j\colon Y \to X$ be the inclusion. Then $j^*\colon H^{2j}(X) \to H^{2j}(Y)$ is an injection by the Weak Lefschetz Theorem (in fact it is an isomorphism if $2j < \dim X - 1$). For any class $z \in H^{2r-2}(X)$ it is true that $\eta_X(\xi \cdot z) = \eta_Y(j^*(z))$. Using this, we have that

$$\eta_X(\xi^{r-2j}x \cdot y) = \eta_Y(j^*(\xi^{r-2j-1}xy)) = \eta_Y(j^*(\xi)^{r-2j-1} \cdot j^*(x) \cdot j^*(y)).$$

By induction, the form $a, b \mapsto \eta_Y(j^*(\xi)^{r-2j-1}a \cdot b)$ on $H^{2j}(Y)$ is positive-definite when j is even, and negative-definite when j is odd. The same can therefore be deduced for our form on $H^{2j}(X)$. $\square$

It is sometimes convenient to have a positive definite form defined on all of $H^{2j}(X;\mathbb{Q})$, not just on the primitive part. We can construct one using the Lefschetz decomposition

$$H^{2j}(X;\mathbb{Q}) = PH^{2j}(X;\mathbb{Q}) \oplus L[PH^{2j-2}(X;\mathbb{Q})] \oplus L^2[PH^{2j-4}(X;\mathbb{Q})] \oplus \cdots$$

in the following way. First, we define the form so that the above summands are orthogonal to each other. Second, the restriction of our form to the summand $L^k[PH^{2j-2k}]$ will be induced by the one from Corollary 3.3, with an appropriate sign thrown in to make it positive definite.

To be more explicit, define the form $\langle -, - \rangle_H$ on $H^{2j}(X;\mathbb{Q})$ by the following formula. If $a = L^i a_0$ and $b = L^k b_0$ where $a_0 \in PH^{2j-2i}$ and $b_0 \in PH^{2j-2k}$, set

$$\langle a, b \rangle_H = \delta_{i,k} \cdot (-1)^{j-i}\eta(L^{r-2i+j+k}a_0 \cdot b_0) = \delta_{i,k} \cdot (-1)^{j-i}\eta(L^{r-2j}a \cdot b).$$

3.4. The Hodge star operator. The above formula for $\langle -, - \rangle_H$ is usually expressed in terms of the Hodge $*$-operator. This is a homomorphism of graded groups $*\colon H^*(X;\mathbb{Q}) \to H^*(X;\mathbb{Q})$ which is specified uniquely by the following properties:

(i) $*^2 = id$
(ii) If $a \in PH^j(X;\mathbb{Q})$ then $*a = (-1)^{\binom{j+1}{2}}L^{r-j}a$.
(iii) If $a \in H^j(X;\mathbb{Q})$ and $j < r-1$ then $L * L(a) = *a$.

(recall that $r = \dim X$).

One can also write an explicit description of the $*$-operator using the Lefschetz decomposition. If $a \in H^j(X;\mathbb{Q})$ and $a = L^i a_i + L^{i+1}a_{i+1} + \cdots$ where $a_i \in PH^{j-2i}(X;\mathbb{Q})$, then

$$*a = (-1)^{\binom{j-2i+1}{2}}L^{r-j+i}a_i + (-1)^{\binom{j-2i+2}{2}}L^{r-j+i+1}a_{i+1} + \cdots$$

Note that $*a$ is almost $L^{r-j}(a)$, except for lots of signs thrown in at different stages of the Lefschetz decomposition.

PROPOSITION 3.5. *For $a, b \in H^j(X;\mathbb{Q})$ one has*

$$\langle a, b\rangle_H = \eta(a \cdot *b).$$

This is a positive definite symmetric bilinear form on $H^j(X;\mathbb{Q})$.

PROOF. We have already explained why $\langle a, b\rangle_H$ is a positive definite symmetric bilinear form—it was constructed in such a way that forces it to be so, using Corollary 3.3. So it is just the first statement of the proposition which must be verified. This is a routine exercise. ☐

EXERCISE 3.6 (Castelnuovo's inequality). Let C and C' be two projective curves over $\mathbb{C}$, and let X be the algebraic surface $C \times C'$. Let $\alpha = [C] \times *$ and $\beta = * \times [C']$, and let $\xi = \alpha + \beta$.

(a) Verify that ξ is a hyperplane section of X.
(b) Let $D \in H^2(X)$, and let $a = \langle D, \alpha\rangle$ and $b = \langle D, \beta\rangle$. Castelnuovo's inequality says

$$\langle D, D\rangle \leq 2ab,$$

with equality only if $D = b\alpha + a\beta$.

To prove this, let $H' = \alpha - \beta$ and check that $H \cdot H' = 0$. Then let $D' = -2D + (a+b)H - (a-b)H'$. Verify that $D \cdot H = 0$, so that $D \in PH^2(X)$. The Hodge Index Theorem says $\langle D', D'\rangle \leq 0$, with equality only when $D' = 0$. Deduce Castelnuovo's result.

[Note: Castelnuovo's result is a basic theorem of algebraic geometry which holds in all characteristics. See [**Ha**, Ex. V.1.9], and also [**G1**].]

4. Hodge theory

Let X be a compact Kähler manifold. Write $H^{p,q}(X) = H^q_{shf}(X, \Omega^p_X)$ for the sheaf cohomology of X with coefficients in the sheaf of holomorphic p-forms. The direct sum $\oplus_{p,q} H^{p,q}(X)$ forms a bi-graded ring in a natural way.

Hodge theory shows that there are natural isomorphisms

$$H^n(X;\mathbb{C}) \cong \oplus_{p+q=n} H^{p,q}(X)$$

which give an isomorphism of graded rings $H^*(X;\mathbb{C}) \cong \oplus_{p,q} H^{p,q}(X)$ (where the latter is graded by total degree).

Hodge further analyzed how Poincaré Duality acts with respect to this decomposition. He proved that if $\dim X = r$ then $H^{p,q}(X) \cong H^{r-p,r-q}(X)$, which in particular shows that $H^{p,q}(X) = 0$ if $p > r$ or $q > r$. The nonzero groups $H^{p,q}(X)$ form the *Hodge diamond*, which we depict in the case of a dimension 4 Kähler manifold:

		$H^{2,2}$		
	$H^{2,1}$		$H^{1,2}$	
$H^{2,0}$		$H^{1,1}$		$H^{0,2}$
	$H^{1,0}$		$H^{0,1}$	
		$H^{0,0}$		

Duality says that the 'antipodal' terms in the Hodge diamond are isomorphic to each other.

Finally, the action of complex conjugation on $H^*(X;\mathbb{C})$ maps $H^{p,q}(X)$ to $H^{q,p}(X)$, thereby showing that these groups are isomorphic. In terms of the picture, this is a reflective symmetry of the Hodge diamond about the central vertical axis. An immediate corollary is that if j is odd then $H^j(X;\mathbb{C})$ is even dimensional.

The numbers $h^{p,q}(X) = \dim H^{p,q}(X)$ are called the **Hodge numbers** of X. We will compute some of these in the examples below.

4.1. Fundamental classes. Let $Z \hookrightarrow X$ be an algebraic subvariety of pure codimension c. It has a fundamental class $[Z] \in H^{2c}(X)$, and Hodge theory shows that its image under $H^{2c}(X) \to H^{2c}(X;\mathbb{C})$ lies purely in the summand $H^{c,c}(X)$. The Hodge conjecture is a converse to this statement:

CONJECTURE 4.2 (Hodge conjecture). *Let X be a smooth, projective, complex variety. If $x \in H^{2c}(X;\mathbb{Q})$ and the image of x under $H^{2c}(X;\mathbb{Q}) \to H^{2c}(X;\mathbb{C})$ lies in $H^{c,c}(X)$, then $x = \sum_i n_i[Z_i]$ for some $n_i \in \mathbb{Q}$ and some algebraic subvarieties $Z_i \hookrightarrow X$ of codimension c.*

REMARK 4.3. Actually, Hodge's original conjecture was made for classes $x \in H^{2c}(X;\mathbb{Z})$ instead of $H^{2c}(X;\mathbb{Q})$; see [**Ho**]. This would have the consequence that every torsion class in $H^{2c}(X;\mathbb{Z})$ was a linear combination of $[Z_i]$'s. But this turned out to be false—a counterexample was given by Atiyah and Hirzebruch [**AH2**, Thm. 6.5], using a construction of Serre's [**Se1**].
[This is not needed in what follows, but here is a brief description of the idea from [**AH2**]. On a smooth scheme, the cohomology fundamental class of an algebraic cycle must survive the Atiyah-Hirzebruch spectral sequence from singular cohomology to complex K-theory. This is essentially because one can build a finite complex of vector bundles resolving the structure sheaf of any algebraic subvariety; the cohomology class of the algebraic cycle survives so that it can 'become' the alternating sum of the vector bundles in the resolution. The differentials in the Atiyah-Hirzebruch spectral sequence are certain cohomology operations, and these then given obstructions for a cohomology class to be algebraic. By analyzing the first k-invariant of BU, one sees that d_3 is an integral lift of Sq^3. If one can find a variety X and a torsion cohomology class on which this operation is nontrivial, then the cohomology class cannot survive the spectral sequence and therefore cannot be algebraic. Serre [**Se1**] proves the remarkable result that for any finite group G and any integer $n > 2$, there is a projective algebraic variety over $\mathbb{C}$ whose homotopy n-type is the same as $K(\mathbb{Z},2) \times K(G,1)$. It is not hard to find a G whose cohomology has a torsion class killed by d_3, and this finishes the counterexample.]

4.4. Compatibility of Hodge and Lefschetz. A hyperplane section of X is, in particular, an algebraic subvariety of codimension 1. So its fundamental class ξ lies in $H^{1,1}(X)$. Since the cup product respects the Hodge decomposition, the Lefschetz operator $L\colon H^j(X;\mathbb{C}) \to H^{j+1}(X;\mathbb{C})$ sends $H^{p,q}(X)$ to $H^{p+1,q+1}(X)$. The Hard Lefschetz theorem then gives that $H^{p,q}(X) \cong H^{r-q,r-p}(X)$. This is another symmetry of the Hodge diamond, this time a reflective symmetry about the central horizontal axis.

Note that the Lefschetz decomposition of $H^*(X;\mathbb{Q})$ into primitive pieces induces a similar decomposition for each of the Hodge groups $H^{p,q}$, because the operator L respects the Hodge bi-grading.

4.5. Sample computations. Now we turn to some examples.

EXAMPLE 4.6 (Projective space). For $\mathbb{C}P^n$ one knows that the cohomology is completely algebraic: the group $H^{2j}(\mathbb{C}P^n)$ is generated by $[\mathbb{C}P^{n-j}]$, which must lie in $H^{j,j}(\mathbb{C}P^n)$. One therefore has that

$$h^{p,q}(\mathbb{C}P^n) = \begin{cases} 1 & \text{if } p = q \text{ and } p \leq n \\ 0 & \text{otherwise.} \end{cases}$$

So the groups in the Hodge diamond for $\mathbb{C}P^n$ are concentrated along the central vertical axis.

EXAMPLE 4.7. (Hodge numbers of complete intersections) For complete intersections the Hodge numbers are completely determined by the degrees of the defining equations. First of all, the Weak Lefschetz theorem shows that the Hodge diamond is the same as for projective space except in the middle dimension. The ranks of the groups in this middle dimension can again be computed via the Hirzebruch-Riemann-Roch theorem. One defines

$$\chi^p(X) = \sum_q (-1)^q \dim H^{p,q}(X;\mathbb{C}) \qquad \text{and} \qquad \chi_y(X) = \sum_p \chi^p(X) y^p.$$

The χ^p's are the Euler characteristics for the rising diagonals in the Hodge diamond. Once one knows these for a complete intersection, the ranks of the groups in the middle dimension can easily be computed.

Suppose $X \hookrightarrow \mathbb{C}P^n$ is a complete intersection defined by a regular sequence of forms having degrees $d_1, d_2, \ldots, d_k$. Hirzebruch's theorem [**H**, Thm. 22.1.1] says that $\chi_y(X)$ is the coefficient of z^n in the formal power series

$$\frac{1}{(1+yz)(1-z)} \cdot \prod_i \left[\frac{(1+yz)^{d_i} - (1-z)^{d_i}}{(1+yz)^{d_i} + y(1-z)^{d_i}} \right]. \tag{4.8}$$

Let's again consider the hypersurface $x^3 + y^3 + z^3 + w^3 + u^3 = 0$ in $\mathbb{C}P^4$. Mathematica can expand the above series for us. The commands to do this (as well as Mathematica's output) are:

```
 In[1]:=   f[z_,y_]:=((1+y*z)^3 - (1-z)^3)/((1+y*z)^3+y*(1-z)^3)
 In[2]:=   g[z_,y_]:=f[z,y]/((1+y*z)(1-z))
 In[3]:=   Series[g[z,y],{z,0,5},{y,0,5}]
Out[3]:=   (3+O[y]^6)z + O[y]^6 z^2 + (1 - 7y +y^2+O[y]^6)z^3 +
                  (1 + 4y - 4y^2 - y^3 + O[y]^6) z^4 +
                    (1 - 2y + 21y^2 - 2y^3 + y^4 + O[y] ^6) z^5
```

For us the relevant information is the coefficient of z^4, which is $1+4y-4y^2-y^3$. From it, we deduce that the Hodge diamond looks as follows:

$$\begin{array}{ccccccc}
 & & & \mathbb{C} & & & \\
 & & 0 & & 0 & & \\
 & 0 & & \mathbb{C} & & 0 & \\
0 & & \mathbb{C}^5 & & \mathbb{C}^5 & & 0 \\
 & 0 & & \mathbb{C} & & 0 & \\
 & & 0 & & 0 & & \\
 & & & \mathbb{C}. & & &
\end{array}$$

EXAMPLE 4.9 (More about hypersurfaces). Let X be a hypersurface of dimension r. The cohomology groups of X (with complex coefficients) consist of $\mathbb{C}$'s in even dimensions, concentrated along the central diagonal in the Hodge diamond, together with the groups $H^{r-i,i}$ (for $0 \leq i \leq r$) on the middle row. Define the **gap** of X to be the number of zeros at the rightmost end of the middle row. That is,

$$\operatorname{gap}(X) = \#\Big\{k \in \big[0, \frac{r}{2}\big] \;\Big|\; H^{k,r-k}(X) = H^{k-1,r-k+1}(X) = \cdots = H^{0,r}(X) = 0\Big\}.$$

The number of groups in the middle row to the right of the central diagonal (including the diagonal itself) is equal to $1 + \frac{r}{2}$ if r is even and $\frac{r+1}{2}$ if r is odd. Define

$$\operatorname{spread}(X) = \begin{cases} 1 + \frac{r}{2} - \operatorname{gap}(X) & \text{if } r \text{ is even,} \\ \frac{r+1}{2} - \operatorname{gap}(X) & \text{if } r \text{ is odd.} \end{cases}$$

Morally, the spread of X is the number of nonzero groups in the middle row which are right of the central diagonal, including the groups on the diagonal itself.

If X is a degree d hypersurface in $\mathbb{C}P^n$, a little work with the power series (4.8) shows that $\operatorname{gap}(X) = \lfloor \frac{n}{d} \rfloor$. So hypersurfaces with small degree have small spread, and the spread increases as the degree increases. Hypersurfaces of degree $n+1$ and higher have full spread (or equivalently, zero gap).

5. Correspondences and the cohomology of manifolds

The previous sections dealt with properties of cohomology which are very particular to compact algebraic varieties over $\mathbb{C}$. The material in this section holds in more generality; it works for spaces which are oriented compact manifolds.

Let X be an oriented, compact manifold of dimension d. The orientation determines an isomorphism $\eta_X \colon H^d(X;\mathbb{Q}) \to \mathbb{Q}$, by sending the cohomology fundamental class of X to 1. The cup product then gives pairings

$$H^i(X;\mathbb{Q}) \otimes H^{d-i}(X;\mathbb{Q}) \to H^d(X;\mathbb{Q}) \cong \mathbb{Q},$$

and a consequence of Poincaré Duality is that these are perfect pairings.

The reader has perhaps seen, in an introductory course on algebraic topology, that the algebra of these perfect pairings can be used to prove the Lefschetz fixed point theorem (see, for example, [**GH**, Chapter 30]). Most textbooks only touch upon these methods, however, without systematically developing the ideas. The goal in this section is precisely to undertake such a systematic development. It will lead us to a nice generalization of the Lefschetz fixed point theorem, and perhaps to a better understanding of it.

Note: The definitions in this chapter lead to several unpleasant signs. It seems like there should be a way to avoid this. These signs are irrelevant for things later in the text, as they disappear when the spaces involved are all even-dimensional (for example, if they are complex algebraic varieties). Still, it seems like there should be an approach to this material which leads to more reasonable signs in the formulas for odd-dimensional manifolds. Perhaps some reader will be inspired to work this out.

5.1. Basic machinery. In this section all cohomology groups have coefficients in $\mathbb{Q}$ (although any field would suffice). Let X be a compact, oriented manifold. We will abuse notation and also use the symbol "X" to denote the dimension of X, in formulas like $H^{X-i}(X) \cong H^i(X)$. Similarly, if $\alpha, \beta \in H^*(X)$ we will also write formulas such as $\alpha\beta = (-1)^{\alpha\beta}\beta\alpha$, where clearly the symbols α and β in the exponent are denoting the dimensions of the corresponding cohomology classes. This abuse of notation is extremely convenient, and in practice there is usually not much chance of confusion. (Actually, there is one chance for confusion: the "$\alpha\beta$" in the exponent might be taken to be the degree of the cohomology class $\alpha\beta$, rather than the product of the degrees. We will never use this interpretation, instead writing $(-1)^{|\alpha\beta|}$ or $(-1)^{\alpha+\beta}$ if necessary).

In our subsequent work in this section we will only use the following properties of singular cohomology.

(1) $H^*(-)$ is a contravariant from spaces to graded-commutative $\mathbb{Q}$-algebras;
(2) $H^*(pt)$ equals $\mathbb{Q}$, concentrated in dimension 0;
(3) $H^i(X) = 0$ for unless $0 \leq i \leq \dim X$;
(4) For each X and Y, the map
$$H^*(X) \otimes H^*(Y) \to H^*(X \times Y), \quad \alpha \otimes \beta \mapsto \pi_1^*(\alpha) \cdot \pi_2^*(\beta)$$
is an isomorphism of rings, where the product on the domain is the graded-tensor product, given by
$$(a \otimes b) \cdot (c \otimes d) = (-1)^{bc}(ac \otimes bd)$$
for homogeneous elements $a, c \in H^*(X)$ and $b, d \in H^*(Y)$. For homogeneous elements $\alpha \in H^*(X)$ and $\beta \in H^*(Y)$ we will often write $\alpha \otimes \beta$ for $\pi_1^*(\alpha) \cdot \pi_2^*(\beta)$, implicitly using the above isomorphism.
(5) For each oriented manifold X there is a chosen isomorphism $\eta_X \colon H^X(X) \to \mathbb{Q}$, and we write Θ_X for the preimage of 1 under this map (this is the cohomological fundamental class of X). We have that $\Theta_{X\times Y} = \Theta_X \otimes \Theta_Y$ for all X and Y.
(6) Finally, for all oriented manifolds X the product maps
$$H^i(X) \otimes H^{X-i}(X) \to H^X(X) \xrightarrow{\eta_X} \mathbb{Q}$$
are perfect pairings.

In Chapter 3 we will want to say that the arguments below work just as well for certain cohomology groups in algebraic geometry. The reason this is true is that our proofs will only use the above properties.

It is useful to extend the map η_X to all of $H^*(X)$ by defining it to be zero on all $H^i(X)$ for $i < X$. For $\alpha, \beta \in H^*(X)$, write
$$\langle \alpha, \beta \rangle = \eta(\alpha \cdot \beta).$$
It is useful to also write $\langle \alpha \rangle$ for $\eta(\alpha)$, as this lets us drop some commas in formulas: e.g., we can write $\langle \alpha, \beta \rangle = \langle \alpha\beta \rangle$.

Poincaré Duality says that $\langle -, - \rangle$ is a perfect pairing $H^i(X) \otimes H^{X-i}(X) \to \mathbb{Q}$. This gives us two canonical isomorphisms
$$\phi_L, \phi_R \colon H^i(X) \xrightarrow{\cong} \operatorname{Hom}_{\mathbb{Q}}(H^{X-i}(X), \mathbb{Q}),$$
given by
$$(\phi_L\alpha)(\beta) = \langle \alpha, \beta \rangle, \qquad (\phi_R\alpha)(\beta) = \langle \beta, \alpha \rangle.$$
These two maps differ by a sign.

Given a map $f\colon X \to Y$, define the cohomology pushforward functor $f_!\colon H^i(X) \to H^{i+Y-X}(Y)$ by the square

$$\begin{tikzcd} H^i(X) \arrow[rr] \arrow[d, "\phi_R"', "\cong"] & & H^{i+Y-X}(Y) \arrow[d, "\phi_R"', "\cong"] \\ \mathrm{Hom}(H^{X-i}(X),\mathbb{Q}) \arrow[rr, "\mathrm{Hom}(f^*,\mathbb{Q})"] & & \mathrm{Hom}(H^{X-i}(Y),\mathbb{Q}). \end{tikzcd}$$

Equivalently, $f_!$ is the unique morphism satisfying the equation

$$\langle \alpha, f_!(\beta)\rangle = \langle f^*(\alpha), \beta\rangle$$

for every $\beta \in H^i(X)$ and $\alpha \in H^{X-i}(Y)$. Intuitively, $f_!$ is the Poincaré dual of the usual pushforward map f_* on homology. But the above approach allows us to construct $f_!$ without ever having to refer to homology at all.

Exercise 5.2. Verify that $\langle xy, z\rangle = \langle x, yz\rangle$ for $x, y, z \in H^*(X)$ and that $\langle f_!(p), q\rangle = (-1)^{q(Y-X)}\langle p, f^*(q)\rangle$ for $p \in H^*(X)$ and $q \in H^*(Y)$. [Remark: Note that the latter formula just doesn't look right, on any level; that is, it doesn't conform to the Koszul sign conventions. This is what we meant by our warning that the signs are sometimes unpleasant. If X and Y are even-dimensional then the sign goes away, and the formula looks more sensible.]

Lemma 5.3 (Projection formula). *Let $f\colon X \to Y$. Then for any $\alpha \in H^i(Y)$ and $\beta \in H^j(X)$, one has $f_!(f^*\alpha \cdot \beta) = \alpha \cdot f_!(\beta)$.*

Proof. This is simply a computation. For any $\gamma \in H^{X-i-j}(Y)$,

$$\begin{aligned} \langle \gamma, f_!(f^*\alpha\cdot\beta)\rangle_Y = \langle f^*\gamma, f^*\alpha\cdot\beta\rangle_X = \langle f^*\gamma\cdot f^*\alpha, \beta\rangle_X &= \langle f^*(\gamma\alpha), \beta\rangle_X \\ &= \langle \gamma\alpha, f_!(\beta)\rangle_Y \\ &= \langle \gamma, \alpha\cdot f_!(\beta)\rangle_Y. \end{aligned}$$

□

Lemma 5.4. *For $\alpha \in H^*(X)$ and $\beta \in H^*(Y)$ one has*

$$(\pi_X^{X\times Y})_!(\alpha\otimes\beta) = \langle\beta\rangle_Y\cdot\alpha \qquad \textit{and} \qquad (\pi_Y^{X\times Y})_!(\alpha\otimes\beta) = (-1)^{X(Y-\beta)}\langle\alpha\rangle_X\cdot\beta.$$

Proof. More computations. For example,

$$\begin{aligned} \big\langle \gamma, (\pi_Y^{X\times Y})_!(\alpha\otimes\beta)\big\rangle_Y = \big\langle (\pi_Y^{X\times Y})^*(\gamma), \alpha\otimes\beta\big\rangle_{X\times Y} &= \big\langle 1\otimes\gamma, \alpha\otimes\beta\big\rangle_{X\times Y} \\ &= \langle (1\otimes\gamma)(\alpha\otimes\beta)\rangle_{X\times Y} \\ &= (-1)^{\gamma\alpha}\eta_{X\times Y}(\alpha\otimes\gamma\beta) \\ &= (-1)^{\gamma\alpha}\langle\alpha\rangle_X\cdot\langle\gamma,\beta\rangle_Y \\ &= (-1)^{(Y-\beta)X}\langle\alpha\rangle_X\cdot\langle\gamma,\beta\rangle_Y. \end{aligned}$$

In the last equality, we can replace the exponent on the -1 because the rest of the expression vanishes unless α has degree X and γ has degree $Y-\beta$.

The other formula is a very similar computation, but even easier. □

If $j\colon X \hookrightarrow Y$ is the inclusion of a submanifold, write $[X] = j_!(1) \in H^{Y-X}(Y)$. This is the "cohomology fundamental class" of X. Also, given $f\colon X \to Y$ let $\Delta_f\colon X \to Y \times X$ be the map $x \mapsto (f(x), x)$. Define $\mathrm{Gr}(f) = f_!(1) \in H^Y(Y\times X)$. This is the fundamental class in $Y \times X$ for the graph of f. (This definition of the graph of f is backwards from what is typically used, e.g. in freshman calculus courses. But the present definition is more consistent with the convention of functions acting on the left, and it will work better with the geometric approach to function composition we will consider below.)

5.5. Correspondences. A correspondence from X to Y is simply a cohomology class $u \in H^*(Y \times X)$. The importance of this concept is that any correspondence induces maps $u^*\colon H^*(Y) \to H^*(X)$ and $u_!\colon H^*(X) \to H^*(Y)$, and in the special case where $u = [\mathrm{Gr}\, f]$ these coincide with f^* and $f_!$. The idea is that a correspondence behaves as if it were a generalized function from X to Y. [Some texts define a correspondence from X to Y to be a cohomology class in $H^*(X \times Y)$ rather than in $H^*(Y \times X)$. As these groups are isomorphic, this is laregly a semantical issue.]

Continuing the analogy with functions, one can define the *composition* of two correspondences. If $v \in H^*(Z \times Y)$ and $u \in H^*(Y \times X)$, define the correspondence $v \circ u \in H^{v+u-Y}(X \times Z)$ by the formula

$$v \circ u = (-1)^{uY+XY}(\pi_{13})_!\big[(\pi_{12})^*(v) \cdot (\pi_{23})^*(u)\big].$$

Here π_{12}, π_{13}, and π_{23} are the evident projections with domain $Z \times Y \times X$ (for example, $\pi_{12}\colon Z \times Y \times X \to Z \times Y$.)

The sign in the above formula is annoying, but it is exactly what is needed to make the composition product associative. For the following result, recall the class $\Delta_X \in H^X(X \times X)$ defined by $\Delta_X = \Delta_!(1)$, where $\Delta\colon X \to X \times X$ is the diagonal embedding.

Proposition 5.6. *Let $w \in H^*(W \times Z)$, $v \in H^*(Z \times Y)$, and $u \in H^*(Y \times X)$. Then*

$$w \circ (v \circ u) = (w \circ v) \circ u.$$

Moreover, $\Delta_Y \circ u = u = u \circ \Delta_X$.

Proof. In this proof we will use several different projection maps. Let us adopt the notation π^{WZX}_{WX} for the projection $W \times Z \times X \to W \times X$, and similarly for other projections. Also, when the domain and range of the projection can be deduced from context we will just write π^*, to simplify the typography. Note that this will sometimes result in several *different* projections all being denoted π^* in the same formula.

The idea of the proof is to manipulate the expression

$$\Omega = (\pi^{WZYX}_{WX})_![\pi^*(w) \cdot \pi^*(v) \cdot \pi^*(u)]$$

in two different ways. One way shows

$$\Omega = \big(\pi^{WZX}_{WZ}\big)_!\Big(\pi^* w \cdot \big(\pi^{ZYX}_{ZX}\big)_!(\pi^* v \cdot \pi^* u)\Big) \tag{5.7}$$

whereas the other way shows

$$\Omega = (-1)^{uZ+XZ}\big(\pi^{WYX}_{WX}\big)_!\Big(\big(\pi^{WZY}_{ZY}\big)_!(\pi^* w \cdot \pi^* v) \cdot \pi^* u\Big). \tag{5.8}$$

Adding signs for the composition products, we see from (5.7) that

$$w \circ (v \circ u) = (-1)^{uY+XY+(u+v-Y)Z+XZ}\Omega.$$

Likewise, equation (5.8) gives that

$$(w \circ v) \circ u = (-1)^{vZ+YZ+uY+XY} \cdot (-1)^{uZ+XZ}\Omega.$$

Comparing these, we find at once that $w \circ (v \circ u) = (w \circ v) \circ u$.

So our task is to prove (5.7) and (5.8). Observe that

$$\begin{aligned}
\Omega &= (\pi^{WZX}_{WX})_!(\pi^{WZYX}_{WZX})_!\Big[(\pi^{WZYX}_{WZX})^*(\pi^*(w))\cdot\big(\pi^*(v)\pi^*(u)\big)\Big]\\
&= (\pi^{WZX}_{WX})_!\Big[\pi^*(w)\cdot(\pi^{WZYX}_{WZX})_!\big(\pi^*(v)\pi^*(u)\big)\Big] \qquad \text{(projection formula)}\\
&= (\pi^{WZX}_{WX})_!\Big[\pi^*(w)\cdot(\pi^{WZYX}_{WZX})_!(\pi^{WZYX}_{ZYX})^*\big(\pi^*(v)\pi^*(u)\big)\Big]\\
&= (\pi^{WZX}_{WX})_!\Big[\pi^*(w)\cdot(\pi^{WZX}_{ZX})^*(\pi^{ZYX}_{ZX})_!\big(\pi^*(v)\pi^*(u)\big)\Big] \qquad \text{(push-pull)}.
\end{aligned}$$

The step labelled "push-pull" uses the identity

$$(\pi^{WZYX}_{WZX})_!(\pi^{WZYX}_{ZYX})^* = (\pi^{WZX}_{ZX})^*(\pi^{ZYX}_{ZX})_!$$

which may be verified by the usual kind of adjointness argument (???).

Likewise, we can start with

$$\Omega = (\pi^{WYX}_{WX})_!(\pi^{WZYX}_{WYX})_!\Big[\big(\pi^*(w)\pi^*(v)\big)\cdot(\pi^{WZYX}_{WYX})^*(\pi^*(u))\Big]$$

and proceed similarly. This time our use of the projection formula comes with a sign: we are looking at something of the form $\pi_!(\alpha\cdot\pi^*(\beta))$, and this equals $(-1)^{\beta\pi}\pi_!(\alpha)\cdot\beta$ (where by convention the dimension of π is the dimension of the codomain minus the dimension of the domain). This is the $(-1)^{uZ}$ sign appearing in (5.8). Later we need to use another push-pull formula, this time

????.

Here the sign comes, ultiimately, from the signs in Lemma 5.4. We will leave the reader to fill in the details here.

IDENTITIES?!!! □

Proposition 5.9. *Let $\alpha\in H^*(X)$, $\beta,\gamma\in H^*(Y)$, and $\delta\in H^*(Z)$. Then*

$$(\delta\otimes\gamma)\circ(\beta\otimes\alpha) = (-1)^{\beta Y}\,\langle\gamma,\beta\rangle\,\cdot\delta\otimes\alpha.$$

Proof. A computation exactly as in Lemma 5.4 shows that

$$\big(\pi^{ZYX}_{ZX}\big)_!\big(p\otimes q\otimes r\big) = (-1)^{(X-r)Y}\langle q\rangle\cdot(p\otimes r).$$

Using this, we compute that

$$\begin{aligned}
(\delta\otimes\gamma)\circ(\beta\otimes\alpha) &= (-1)^{(\alpha+\beta)Y+XY}\cdot\big(\pi^{ZYX}_{ZX}\big)_!\big((\delta\otimes\gamma\otimes 1)\cdot(1\otimes\beta\otimes\alpha)\big)\\
&= (-1)^{(\alpha+\beta)Y+XY}\cdot\big(\pi^{ZYX}_{ZX}\big)_!\big(\delta\otimes(\gamma\beta)\otimes\alpha\big)\\
&= (-1)^{(\alpha+\beta)Y+XY}\cdot(-1)^{(X-\alpha)Y}\langle\gamma,\beta\rangle\cdot(\delta\otimes\alpha)\\
&= (-1)^{\beta Y}\cdot\langle\gamma,\beta\rangle\cdot(\delta\otimes\alpha).
\end{aligned}$$

□

We now define the functions u^* and $u_!$ induced by a correspondence. Write π_1 and π_2 for the projections from $Y\times X$ to Y and X. Let $u\in H^*(Y\times X)$. For $\alpha\in H^*(Y)$ and $\beta\in H^*(X)$, define

$$u^*(\alpha)=\alpha\circ u \qquad\text{and}\qquad u_!(\beta)=u\circ\beta.$$

Here α is identified with an element of $H^*(pt\times Y)$ in the first equation, and β is identified with an element of $H^*(X\times pt)$ in the second.

The above formulas define maps $u^*\colon H^*(Y) \to H^{*+u-Y}(X)$ and $u_!\colon H^*(X) \to H^{*+v+Y-X}(Y)$. If $v \in H^*(Z \times Y)$ and $u \in H^*(Y \times X)$, note that

$$(v \circ u)^*(\alpha) = u^*(v^*(\alpha)) \qquad \text{and} \qquad (v \circ u)_!(\beta) = u_!(v_!(\beta))$$

for $\alpha \in H^*(Z)$ and $\beta \in H^*(X)$. These formulas are direct consequences of the associativity of the composition product, Proposition 5.6.

Remark 5.10. By using the definition of the composition product, we can write

$$u^*(\alpha) = (-1)^{uY+XY}(\pi_2)_!((\pi_1)^*(\alpha) \cdot u) \in H^{\alpha+u-Y}(X)$$

and

$$u_!(\beta) = (-1)^{\beta X}(\pi_1)_!(u \cdot (\pi_2)^*\beta) \in H^{\beta+u-X}(Y)$$

for $u \in H^*(Y \times X)$, $\alpha \in H^*(Y)$, and $\beta \in H^*(X)$. As a consequence, one obtains the following formulas as well:

$$\begin{aligned}\langle \beta, u^*(\alpha)\rangle &= (-1)^{uY+XY}\langle \pi_2^*(\beta) \cdot \pi_1^*(\alpha) \cdot u\rangle_{X\times Y}\\ \langle \alpha, u_!(\beta)\rangle &= (-1)^{\beta X}\langle \pi_1^*(\alpha) \cdot u \cdot \pi_2^*(\beta)\rangle_{X\times Y}.\end{aligned}$$

Lemma 5.11. *Let $f\colon X \to Y$. Then for $\alpha \in H^*(Y)$ and $\beta \in H^*(X)$ one has*

$$[\mathrm{Gr}\, f]^*(\alpha) = (-1)^{XY+Y} f^*(\alpha) \qquad \textit{and} \qquad [\mathrm{Gr}\, f]_!(\beta) = (-1)^{\beta(X+Y)} f_!(\beta).$$

Proof. Let $\alpha \in H^i(Y)$, and let $\beta \in H^{X-i}(X)$. Then

$$\begin{aligned}(-1)^{Y^2+XY}\langle \beta, [\mathrm{Gr}\, f]^*(\alpha)\rangle &= \langle \pi_2^*(\beta) \cdot \pi_1^*(\alpha) \cdot [\mathrm{Gr}\, f]\rangle\\ &= \langle \pi_2^*(\beta) \cdot \pi_1^*(\alpha), \tilde{f}_!(1)\rangle\\ &= \langle \tilde{f}^*(\pi_2^*(\beta) \cdot \pi_1^*(\alpha)), 1\rangle\\ &= \langle \beta \cdot f^*(\alpha), 1\rangle\\ &= \langle \beta, f^*(\alpha)\rangle.\end{aligned}$$

This shows $[\mathrm{Gr}\, f]^*(\alpha) = (-1)^{Y+XY} f^*\alpha$ (using that $Y \equiv Y^2$ mod 2), and a similar argument shows the other identity. $\square$

The signs in the above lemma are annoying, but I don't know how to avoid them. Note that they disappear if f is a map $X \to X$, and of course they also disappear if both X and Y are even-dimensional.

Exercise 5.12. Let $u \in H^*(Y \times X)$, $a \in H^*(X)$, and $b \in H^*(Y)$. Verify that

$$\langle b, u_!(a)\rangle = \langle u^*(b), a\rangle \cdot (-1)^{aY+aX+uY+XY}.$$

5.13. Correspondences and cohomology homomorphisms. Let X and Y be oriented manifolds. A homomorphism of vector spaces $h\colon H^*(Y) \to H^*(X)$ is said to be **homogeneous of degree** c if $h(H^i(Y)) \subseteq H^{i+c}(X)$ for all $i \in \mathbb{Z}$. Write $\mathrm{Hom}^c(H^*(Y), H^*(X))$ for the vector space consisting of all such homomorphisms, and write

$$\mathrm{Hom}(H^*(Y), H^*(X)) = \bigoplus_{c\in\mathbb{Z}} \mathrm{Hom}^c(H^*(Y), H^*(X)).$$

We have $\mathbb{Q}$-linear maps

$$H^i(Y \times X) \to \mathrm{Hom}^{i-Y}(H^*(Y), H^*(X)), \qquad u \mapsto u^*$$

which we may regard as a degree $-Y$ map of graded vector spaces

$$H^*(Y \times X) \longrightarrow \mathrm{Hom}(H^*(Y), H^*(X)).$$

Lemma 5.14. *Let $\alpha \in H^*(X)$ and $\beta \in H^*(Y)$. Then for $w \in H^*(X)$ one has*
$$(\beta \otimes \alpha)^*(w) = (-1)^{\beta}\langle \beta, w\rangle \cdot \alpha.$$
In particular, note that $(\beta \otimes \alpha)^*$ is nonzero only on $H^{Y-\beta}(Y)$, and that its image equals the subspace of $H^{\alpha}(X)$ generated by α.

Proof. This is actually a corollary of Lemma 5.9, because
$$\begin{aligned}(\beta \otimes \alpha)^*(w) = w \circ (\beta \otimes \alpha) = (1 \otimes w) \circ (\beta \otimes \alpha) &= (-1)^{\beta Y}\langle w, \beta\rangle(1 \otimes \alpha)\\ &= (-1)^{\beta Y + \beta w}\langle \beta, w\rangle \cdot \alpha.\end{aligned}$$
But $\langle \beta, w\rangle$ is nonzero only when $|w| = Y - |\beta|$, and so we can write
$$(-1)^{\beta Y + \beta w} = (-1)^{\beta(Y+w)} = (-1)^{\beta^2} = (-1)^{\beta}.$$
□

Corollary 5.15. *Suppose $u \in H^i(Y) \otimes H^j(X) \subseteq H^{i+j}(Y \times X)$. Then the map $u^*\colon H^*(Y) \to H^*(X)$ is nonzero only on $H^{Y-i}(Y)$ and its image is contained in $H^j(X)$.*

Proof. Simply write $u = \sum_k \beta_k \otimes \alpha_k$ and use Lemma 5.14. □

Proposition 5.16. *The map $\Gamma\colon H^*(Y \times X) \to \mathrm{Hom}(H^*(Y), H^*(X))$ is an isomorphism of graded vector spaces.*

Proof. Given a homogeneous map $h\colon H^*(Y) \to H^*(X)$, we write $h = \sum_k h_k$ where h_k is nonzero only on $H^k(Y)$ (and equals the restriction of h thereon). To prove that Γ is surjective it will be sufficient to show that each h_k is in the image. So without loss of generality, replace h by h_k.

Now h is a map $H^k(Y) \to H^j(X)$. Pick a basis $\alpha_1, \ldots, \alpha_p$ for $H^j(X)$. Then one obtains unique functionals $\phi_1, \ldots, \phi_p$ on $H^k(Y)$ such that
$$h(u) = \sum_s \phi_s(u)\alpha_s.$$
Using the nondegenerate pairing $H^k(Y) \otimes H^{Y-k}(Y) \to \mathbb{Q}$, there exist unique $\beta_s \in H^{Y-k}(Y)$ such that $\phi_s = \langle \beta_s, -\rangle$. It is now clear from Lemma 5.14 that
$$h = \Gamma\Big((-1)^{(Y-k)} \cdot \sum_s \beta_s \otimes \alpha_s\Big).$$

This proved the surjectivity of Γ. The injectivity is an immediate consequence of the two words "unique" appearing in the previous paragraph. □

5.17. The transpose operator. Given spaces X and Y, let $t_{X,Y}\colon X \times Y \to Y \times X$ be the usual twist map. We will usually abbreviate $t_{X,Y} = t$.

Lemma 5.18. *For $\beta \in H^*(Y)$ and $\alpha \in H^*(X)$ one has $t^*(\beta \otimes \alpha) = (-1)^{\alpha\beta}\alpha \otimes \beta$ and $t_!(\alpha \otimes \beta) = (-1)^{\alpha\beta + XY}\beta \otimes \alpha$.*

Proof. ??? The second identity follows immediately from the first, using adjointness (i.e., compute $\langle p \otimes q, t_!(\alpha \otimes \beta)\rangle$). □

For $u \in H^*(Y \times X)$ we define $u^t = t^*(u)$ and call this the **transpose** of u.

Proposition 5.19. *Let $u \in H^*(Y \times X)$ and $v \in H^*(Z \times Y)$. Then*
$$(v \circ u)^t = (-1)^{uY + vY + uv}(u^t \circ v^t).$$

PROOF. Let $\sigma\colon X\times Y\times Z \to Z\times Y\times X$ be the evident map. We first observe that $\sigma_!\sigma^* = (-1)^{XY+YZ+XZ}\cdot id$. Next we compute:

$$\begin{aligned}(v\circ u)^t &= (-1)^{uY+XY}(t^*_{X,Z})\big(\pi^{ZYX}_{ZX}\big)_!\Big((\pi^{ZYX}_{ZY})^*(v)\cdot(\pi^{ZYX}_{YX})^*(u)\Big)\\ &= (-1)^{uY+XY+XZ}(t_{X,Z})_!\big(\pi^{ZYX}_{ZX}\big)_!\Big((\pi^{ZYX}_{ZY})^*(v)\cdot(\pi^{ZYX}_{YX})^*(u)\Big)\\ &= (-1)^{uY+YZ}\big(\pi^{ZYX}_{XZ}\big)_!\sigma_!\sigma^*\Big((\pi^{ZYX}_{ZY})^*(v)\cdot(\pi^{ZYX}_{YX})^*(u)\Big)\\ &= (-1)^{uY+YZ}\big(\pi^{XYZ}_{XZ}\big)_!\Big((\pi^{XYZ}_{ZY})^*(v)\cdot(\pi^{XYZ}_{YX})^*(u)\Big)\\ &= (-1)^{uY+YZ}\big(\pi^{XYZ}_{XZ}\big)_!\Big((\pi^{XYZ}_{YZ})^*(v^t)\cdot(\pi^{XYZ}_{XY})^*(u^t)\Big)\\ &= (-1)^{uY+YZ+uv}\big(\pi^{XYZ}_{XZ}\big)_!\Big((\pi^{XYZ}_{XY})^*(u^t)\cdot(\pi^{XYZ}_{YZ})^*(v^t)\Big)\\ &= (-1)^{uY+YZ+uv}\cdot(-1)^{vY+YZ}\cdot(u^t\circ v^t)\\ &= (-1)^{uY+vY+uv}(u^t\circ v^t).\end{aligned}$$

$\square$

EXERCISE 5.20. Verify that $(u^t)^*(\alpha) = (-1)^{uX+\alpha X+u\alpha}\cdot u_!(\alpha)$ for $\alpha\in H^*(X)$ and $u\in H^*(Y\times X)$. Also verify the formula

$$\langle (u^t)^*(\alpha),\beta\rangle = \langle a, u^*(\beta)\rangle\cdot(-1)^{uY+XY+uX+bX+bu}.$$

LEMMA 5.21. *If X is an oriented manifold then $(\Delta_X)^t = (-1)^X\Delta_X$ in $H^*(X\times X)$.*

PROOF. $t^*(\Delta_X) = (-1)^{X^2}t_!(\Delta_X) = (-1)^X\cdot t_!(\Delta_!(1)) = (-1)^X\cdot\Delta_!(1)$. $\square$

5.22. The Lefschetz trace formulas. If $w\in H^X(X\times X)$ then $w^*\colon H^*(X)\to H^*(X)$ has degree 0. Let us write $\mathrm{Tr}_i(w^*)$ for the trace of $w^*|_{H^i(X)}$. More generally, for any $w\in H^*(X\times X)$ let us write $\mathrm{Tr}_i(w^*)$ for the trace of $\pi_i\circ w^*\circ\pi_i$, where π_i is the projection of $H^*(X)$ onto $H^i(X)$. Note that with this definition, $\mathrm{Tr}_i(w^*)$ is simply the trace of $(w_X)^*$ where w_X is the component of w in $H^X(X\times X)$.

The following result gives four versions of a generalized Lefschetz trace formula. This is the main result we have been aiming for in this section.

THEOREM 5.23. *Let $u\in H^*(Y\times X)$ and $v\in H^*(X\times Y)$, so that $v\circ u\in H^*(X\times X)$. Then one has*

$$\sum_k(-1)^k\,\mathrm{Tr}_k(v\circ u)^* = (-1)^{X+XY+uY}\cdot\langle v^t,u\rangle_{Y\times X} = (-1)^{(X+1)(Y-u)}\langle u,v^t\rangle_{Y\times X}$$

and also

$$\sum_k(-1)^k\,\mathrm{Tr}_k(v\circ u)^* = (-1)^{X+u+uX}\cdot\langle u^t,v\rangle_{X\times Y} = (-1)^{X+uY}\langle v,u^t\rangle_{X\times Y}$$

Before giving the proof, we note a few consequences. First observe that if $u\in H^X(X\times X)$ and we take $v=\Delta_X$, then the first formula gives

$$\sum_k(-1)^k\,\mathrm{Tr}_k(u^*) = \langle\Delta,u\rangle_{X\times X}$$

(where we have used the fact that $\Delta^t = (-1)^X\Delta$). This is the classical Lefschetz fixed point formula, usually given when $u=[\mathrm{Gr}\,f]$ for some map $f\colon X\to X$.

If $u \in H^*(Y \times X)$, write $u = \sum_i u_i$ where $u_i \in H^{Y-i}(Y) \otimes H^{u+i-Y}(X)$. Note that u_i^* is nonzero only on $H^i(Y)$, and is just the restriction of u^* on this subspace. When $u = \Delta \in H^X(X \times X)$, then $(\Delta_i)^*$ is simply the projection from $H^*(X)$ onto $H^i(X)$.

Corollary 5.24. *Let $u \in H^X(X \times X)$. Then*

$$\mathrm{Tr}_i(u^*) = (-1)^i \cdot \langle \Delta_{X-i}, u\rangle.$$

Proof. Since $\Delta_X^t = (-1)^X \Delta_X$, it follows readily that $\Delta_i^t = (-1)^X \Delta_{X-i}$. Now just apply one of the Lefschetz trace formulas:

$$(-1)^i \,\mathrm{Tr}_i(u^*) = \sum_k (-1)^k \,\mathrm{Tr}_k(u \circ \Delta_i)^* = (-1)^X \cdot \langle \Delta_i^t, u\rangle = \langle \Delta_{X-i}, u\rangle.$$

□

Let $u \in H^X(X \times X)$, so that u^* is a collection of maps $H^i(X) \to H^i(X)$. Suppose that instead of just wanted the trace of u^* we wanted the characteristic polynomials. By linear algebra, the coefficients of the characteristic polynomial can be computed from the traces of the iterates of u^* (see the proof below for more details about this). This leads to the following:

Proposition 5.25. *Fix i, and let let $n = \dim_{\mathbb{Q}} H^i(X)$. Let $u \in H^X(X \times X)$, and let $p(t) = \det(u^*|_{H^i(X)} - tI)$ be the characteristic polynomial of u^* acting on $H^i(X)$. Then the coefficients of $p(t)$ are obtained as universal rational algebraic expressions of the numbers*

$$\langle \Delta_{X-i}, u\rangle,\ \langle \Delta_{X-i}, u \circ u\rangle,\ \cdots,\ \langle \Delta_{X-i}, u^{\circ(n)}\rangle.$$

Proof. Let $\lambda_1, \ldots, \lambda_n$ be the complex eigenvalues of u^* acting on $H^i(X)$. Then the coefficients of $p(t)$ are the symmetric functions on the λ's, and these can be written in terms of the Newton polynomials in the power sums $\lambda_1^k + \cdots + \lambda_n^k$. Yet this power sum is nothing other than

$$\mathrm{tr}\Big((u^*)^{\circ k}|_{H^i(X)}\Big) = \mathrm{tr}\Big((u^{\circ k})^*|_{H^i(X)}\Big) = (-1)^i \langle \Delta_{X-i}, u^{\circ(k)}\rangle.$$

□

This is as far as we need to go in the present chapter. We close with the proof of the trace formulas:

Proof of Theorem 5.23. This is yet another computation. We will only prove the first formula, the others being very similar.

First note that we may assume $|u| + |v| = X + Y$, otherwise both sides of the equation are zero. Next write $u = \sum u_i$ and $v = \sum v_j$ with $u_i \in H^{u-i}(Y) \otimes H^i(X)$ and $v_j \in H^{X-j}(X) \otimes H^{Y+j-u}(Y)$. Then $v \circ u = \sum_{i,j}(v_j \circ u_i)$ and $\langle v^t, u\rangle = \sum_{i,j}\langle v_j^t, u_i\rangle$. It is easy to see that $v_j \circ u_i = 0$ unless $j = i$, and likewise $\langle v_j^t \cdot u_i\rangle = 0$ unless $j = i$. So we may reduce to the case $u = u_i$ and $v = v_i$; that is, $u \in H^{u-i}(Y) \otimes H^i(X)$ and $v \in H^{X-i}(X) \otimes H^{Y+i-u}(Y)$. Note that in this case v^* is a map $H^i(X) \to H^{Y+i-u}(Y)$ and u^* is a map $H^{Y-u+i}(Y) \to H^i(X)$, so $(v \circ u)^* = u^* \circ v^*$ is a map $H^i(X) \to H^i(X)$. Thus, $\sum_k (-1)^k \,\mathrm{Tr}_k(v \circ u)^* = (-1)^i \,\mathrm{Tr}_i(v \circ u)^*$.

Let $\{\beta_s\}$ be any basis for $H^i(X)$, and let $\{\hat{\beta}_s\}$ be the dual basis for $H^{X-i}(X)$ defined by $\langle \beta_s, \hat{\beta}_t\rangle = \delta_{s,t}$. We may write

$$u = \sum \alpha_s \otimes \beta_s \qquad \text{and} \qquad v = \sum \hat{\beta}_s \otimes \gamma_s$$

for unique $\alpha_s \in H^{u-i}(Y)$ and $\gamma_s \in H^{Y-u+i}(Y)$. Then

$$v \circ u = \sum_{s,t} (\hat{\beta}_s \otimes \gamma_s) \circ (\alpha_t \otimes \beta_t) = \sum_{s,t} (-1)^{\alpha Y} \langle \gamma_s, \alpha_t \rangle (\hat{\beta}_s \otimes \beta_t)$$

Now

$$\begin{aligned} (v \circ u)^*(\beta_r) &= (-1)^{\alpha Y} \sum_{s,t} \langle \gamma_s, \alpha_t \rangle \cdot (\hat{\beta}_s \otimes \beta_t)^*(\beta_r) \\ &= (-1)^{\alpha Y + \hat{\beta}} \sum_{s,t} \langle \gamma_s, \alpha_t \rangle \cdot \langle \hat{\beta}_s, \beta_r \rangle \cdot \beta_t \\ &= (-1)^{\alpha Y + \hat{\beta} + \beta\hat{\beta}} \sum_{t} \langle \gamma_r, \alpha_t \rangle \cdot \beta_t. \end{aligned}$$

Therefore

$$\mathrm{Tr}(v \circ u)^* = (-1)^{\alpha Y + \hat{\beta} + \beta\hat{\beta}} \sum_r \langle \gamma_r, \alpha_r \rangle.$$

Likewise, we can also compute

$$\begin{aligned} v^t \cdot u = \sum_{s,t} (-1)^{\hat{\beta}\gamma} (\gamma_s \otimes \hat{\beta}_s) \cdot (\alpha_t \otimes \beta_t) &= \sum_{s,t} (-1)^{(\hat{\beta}\gamma + \hat{\beta}\alpha)} (\gamma_s \alpha_t \otimes \hat{\beta}_s \beta_t) \\ &= (-1)^{\hat{\beta} Y} \sum_{s,t} (\gamma_s \alpha_t \otimes \hat{\beta}_s \beta_t). \end{aligned}$$

Therefore

$$\langle v^t, u \rangle = (-1)^{\hat{\beta} Y + \beta\hat{\beta}} \sum_s \langle \gamma_s, \alpha_s \rangle.$$

Comparing our formulas for $\mathrm{Tr}(v \circ u)^*$ and $\langle v^t, u \rangle$, we find they differ by the sign $(-1)^S$ where

$$S = \alpha Y + \hat{\beta} + \hat{\beta} Y = (u - i)Y + (X - i) + (X - i)Y.$$

Since $(-1)^S = (-1)^{X + XY + uY + i}$, this completes our proof.

$\square$

Looking ahead.

For oriented manifolds, cup product on cohomology is dual to the intersection product in homology. The former is easier to define, and easier to work with, but it is the latter which gives us connections to geometry. The results in this section can be thought of as giving a geometric interpretation of the coefficients of the characteristic polynomial of u^* acting on $H^i(X)$: the coefficients can be understood in terms of the intersection products of the duals of $u^{\circ(k)}$ and Δ_{X-i}. Looking ahead to the next section, where we return to the Weil conjectures, the importance of this observation is as follows. It in some sense shows that the technology needed to prove the Weil conjectures, while ostensibly requiring a well-developed cohomology theory for algebraic varieties, can be pared down further and further until it just involves producing certain algebraic cycles and studying their intersection numbers. This is the main idea behind Grothendieck's so-called "Standard Conjectures".

CHAPTER 3

A second look at the Weil conjectures

The Standard Conjectures on algebraic cycles were developed independently by Grothendieck and Bombieri in the mid 1960s, in an effort to better explain the Weil conjectures (chiefly the Riemann hypothesis). In the literature one finds a brief expository outline by Grothendieck [**G2**] as well as two detailed treatments by Kleiman [**Kl1, Kl2**].

Before describing the conjectures, let us briefly recall the setting at that time. For varieties over an algebraically closed field k, Grothendieck and his collaborators had defined a family of cohomology theories $X \mapsto H^*(X;\mathbb{Q}_l)$, one for every prime l different from the characteristic of k. These theories satisfied Poincaré Duality, a Lefschetz trace formula, as well as many other nice properties. If X was defined over a finite field $\mathbb{F}_q$, one obtained a rational expression

$$Z(X,t) = \frac{P_1(t)P_3(t)\cdots P_{2d-1}(t)}{P_0(t)P_2(t)\cdots P_{2d}(t)}$$

where $P_i(t)$ is the characteristic polynomial of F^* acting on $H^i(\bar{X};\mathbb{Q}_l)$. Here $\bar{X} = X_{\overline{\mathbb{F}}_q}$ and $F\colon \bar{X} \to \bar{X}$ is the geometric Frobenius map. Poincaré Duality yielded a functional equation for the $Z(X,t)$. Note that the $P_i(t)$'s are polynomials with coefficients in $\mathbb{Q}_l$. The following two things were at that time conjectured but not proven:

(1) (Independence of l) The coefficients of the $P_i(t)$'s are integers, and are independent of l.
(2) (Riemann hypothesis) The reciprocal roots of $P_i(t)$ have absolute value $q^{i/2}$.

The primary goal of the Standard Conjectures was to show how these claims would follow from more fundamental assertions about algebraic cycles and the behavior of cohomology theories.

Let X be a smooth, connected, projective variety defined over a field k. Recall that one has the Chow groups $CH^*(X)$, and that these come equipped with a multiplication induced by intersection of cycles. Let $d = \dim X$. If $U \hookrightarrow X$ and $W \hookrightarrow X$ are subvarieties such that $\dim U + \dim W = \dim X$, then one can move W in its rational equivalence class so that $U \cap W$ is a finite set of closed points. For each point $p \in U \cap W$ one has an intersection multiplicity defined by

$$i(U,W;p) = \sum_{i=0}^{\infty} (-1)^i \dim_k \operatorname{Tor}_i^{\mathcal{O}_{X,p}}(\mathcal{O}_{U,p},\mathcal{O}_{W,p}).$$

It is known that this number is positive, and of course it is an integer. If $U \cap W = \{p_1,\ldots,p_s\}$, then the product of $[U]$ and $[W]$ in $\mathrm{CH}^*(X)$ is

$$[U]\cdot[W] = \sum_j i(U,W;p_j)[p_j] \in \mathrm{CH}^d(X).$$

But recall that $\mathrm{CH}^d(X) \cong \mathbb{Z}$, and the isomorphism sends any $[p]$ to 1. So we can write

$$[U]\cdot[W] = \left(\sum_j i(U, W; p_j)\right)[*]$$

for any closed point $* \in X$.

The first point of the Standard Conjectures will be to interpret the coefficients of the polynomials $P_i(t)$ in terms of intersection products of certain algebraic cycles, independent of any cohomology theory. This will establish the independence of l.

For the Riemann hypothesis one has to work a bit harder. Essentially the idea is the following. Consider the map $F^*\colon H^*(\bar{X};\mathbb{Q}_l) \to H^*(\bar{X};\mathbb{Q}_l)$, and let $\xi \in H^2(\bar{X};\mathbb{Q}_l)$ be the class of a hyperplane section. It is easy to establish that $F^*(\xi) = q\xi$. Define a new map $\phi\colon H^*(\bar{X};\mathbb{Q}_l) \to H^*(\bar{X};\mathbb{Q}_l)$ by letting it act on elements $x \in H^n(\bar{X};\mathbb{Q}_l)$ by

$$\phi(x) = \frac{F^*(x)}{q^{n/2}}.$$

This is still a ring map, and it now satisfies $\phi(\xi) = \xi$. So it commutes with the Lefschetz operator $L(x) = x\cdot\xi$. If one postulates that the Hard Lefschetz theorem is true for $H^*(\bar{X};\mathbb{Q}_l)$, then there is an associated Lefschetz decomposition into primitive pieces. The map ϕ will respect this decomposition. On the primitive component $PH^{2i}(\bar{X};\mathbb{Q}_l)$ one can look at the symmetric bilinear form $\langle x, y\rangle = (-1)^i\eta(L^?x\cdot y)$. The map ϕ preserves this form. If the form is positive-definite in some sense (note that the term doesn't quite make sense since our field is $\mathbb{Q}_l$), then it will be unitary after complexification. So the complexified ϕ will be a unitary operator and hence its eigenvalues will have norm 1. This is equivalent to the Riemann hypothesis, that the eigenvalues of F^* on $H^n(\bar{X};\mathbb{Q}_l)$ have norm $q^{n/2}$.

Now, as we mentioned the above paragraph does not quite make sense because one cannot talk about a positive-definite form over $\mathbb{Q}_l$. So the argument has to be done a bit differently. Still, the above paragraph gives the general idea.

This line of argument, which ties the Riemann hypothesis to the positivity of a certain bilinear form, has a history which is worth recounting. Weil proved the Riemann hypothesis for curves by using Castelnuovo's inequality for the intersection product of curves on an algebraic surface—and this is exactly a positivity result about the intersection form. Weil explained in [**W6**] that Castelnuovo's inequality is—for complex varieties—a version of the Hodge Index Theorem for surfaces. Speculating that the general Riemann Hypothesis might follow from the general Hodge Index Theorem, Weil asked whether one could prove an analog of the Riemann Hypothesis for Kähler manifolds. This challenge was taken up by Serre, who answered it in [**Se4**] by giving the argument from two paragraphs back (except using singular cohomology rather than étale cohomology, where the argument actually makes sense because the coefficients are in $\mathbb{Q}$ rather than $\mathbb{Q}_l$).

1. Weil cohomology theories

We follow [**Kl2**] in making the following definition. Fix an algebraically closed field k, and a characteristic zero field E. A **Weil cohomology theory** with coefficient field E is a contravariant functor $X \mapsto H^*(X)$ from smooth, connected, projective k-schemes to graded-commutative E-algebras satisfying the following properties:

(1) Each $H^i(X)$ is finite-dimensional, and nonzero only in the range $0 \leq i \leq 2\dim X$.
(2) For varieties X of dimension r, there is a functorial isomorphism $\eta\colon H^{2r}(X) \to E$. For each $0 \leq i \leq 2r$ the cup product
$$H^i(X) \otimes H^{2r-i}(X) \to H^{2r}(X)$$
is a perfect pairing.
(3) For each X and Y, the map
$$H^*(X) \otimes H^*(Y) \to H^*(X \times Y)$$
induced by the projections $X \times Y \to X$ and $X \times Y \to Y$ is an isomorphism.
(4) There is a natural map $\gamma_X : Z^i(X) \to H^{2i}(X)$ satisfying ???
(5) (Weak Lefschetz) If $X \hookrightarrow W$ is the inclusion of a smooth hyperplane section, then $H^i(W) \to H^i(X)$ is an isomorphism for $i < \dim X$ and an injection for $i = \dim X$.
(6) (Hard Lefschetz) Let $\xi \in H^2(X)$ be the cycle class of a smooth hyperplane section of X, and let $L\colon H^*(X) \to H^*(X)$ be given by $L(x) = x \cdot \xi$. Then for any $i \leq \dim X$, the map
$$L^{r-i}\colon H^i(X) \to H^{2r-i}(X)$$
is an isomorphism.

Let $A^i(X) \subseteq H^{2i}(X)$ be the rational vector space spanned by the cycle classes of codimension i algebraic subvarieties. Note that it is not at all clear that $A^i(X)$ is finite-dimensional. We will say that a class $x \in H^{2i}(X)$ is **algebraic** (or *rationally algebraic*) if it lies in $A^i(X)$. Likewise, x is **integrally algebraic** if it lies in the $\mathbb{Z}$-submodule generated by the fundamental classes of algebraic subvarieties.

Using the Künneth isomorphism $H^*(X \times Y) \cong H^*(X) \otimes H^*(Y)$, a class $x \in H^k(X \times Y)$ will decompose as
$$x = \sum_{i=0}^{k} x_i' \otimes x_{k-i}'' \in \bigoplus_i H^i(X) \otimes H^{k-i}(Y).$$
We will write $x_i = x_i' \otimes x_{k-i}''$, so that $x = \sum_i x_i$. The class x_i is called the ***i*th Künneth component** of x.

If X has dimension r, let $\Delta \in H^{2r}(X \times X)$ be the cycle class of the diagonal. The following is the geometric version of the Lefschetz trace formula:

THEOREM 1.1. *Let $u \in H^{2r}(X \times X)$, and write $u^*\colon H^*(X) \to H^*(X)$ for the resulting map. Then*
(a) $\operatorname{tr}\big(u^*|_{H^i(X)}\big) = (-1)^i \eta(u \cdot \pi^{2r-i})$.
(b) $\sum_{i=0}^{2r} (-1)^i \operatorname{tr}\big(u^*|_{H^i(X)}\big) = \eta(u \cdot \Delta)$.

2. The Künneth conjecture

Consider the following statement:

$\mathcal{K}\mathrm{u}(X)$: For each i, the Künneth component π_i of the identity is rationally algebraic, and comes from an element of $Z^i(X) \otimes \mathbb{Q}$ which is independent of the Weil cohomology theory.

PROPOSITION 2.1. *Suppose that* $\mathcal{K}\mathrm{u}(X)$ *holds, for a given variety* X *over our ground field* k. *Then for any map* $f\colon X \to X$, *the characteristic polynomial for* $f^*\colon H^i(X) \to H^i(X)$ *has integral coefficients which are independent of the cohomology theory* H^*.

Before proving this we establish some algebraic lemmas. In the following lemma, we mostly care about part (b). But we offer part (a) as a special case which helps understand what is going on in the proof of (b).

LEMMA 2.2. *Let* $m \geq 1$ *be an integer.*

(a) If $u \in \mathbb{Q}$ *satisfies* $u^k \in \frac{1}{m}\mathbb{Z}$ *for all* $k \geq 1$, *then* $u \in \mathbb{Z}$.

(b) Suppose $\lambda_1, \dots, \lambda_d \in \mathbb{C}$ *are such that* $\lambda_1^k + \lambda_2^k + \cdots + \lambda_d^k \in \frac{1}{m}\mathbb{Z}$ *for all* $k \geq 1$. *Then the* λ_i*'s are all algebraic integers. (In fact this holds not just for* $\lambda_i \in \mathbb{C}$ *but for elements of any extension field of* $\mathbb{Q}$*).*

PROOF. For part (a), consider the subring $\mathbb{Z}[u] \subseteq \mathbb{Q}$. The hypothesis is that $\mathbb{Z}[u] \subseteq \frac{1}{m}\mathbb{Z}$, and hence $\mathbb{Z}[u]$ is finitely-genereated as a $\mathbb{Z}$-module (because the same is true of $\frac{1}{m}\mathbb{Z}$). It follows that $\mathbb{Z} \hookrightarrow \mathbb{Z}[u]$ is an integral extension of rings, but of course the only rational numbers which are integral over $\mathbb{Z}$ are the integers themselves.

The above proof is a little bit like cracking a walnut with a sledgehammer, so we also offer the following elementary argument. Write $u = \frac{k}{m}$, and let $d = (k, m)$. Write $k = dk_1$ and $m = dm_1$. The only way that one could have $u^2 \in \frac{1}{m}\mathbb{Z}$ is if $m_1 | dk_1^2$. As m_1 and k_1 are relatively prime, this means $m_1 | d$. Likewise, the only way that one could have $u^r \in \frac{1}{m}\mathbb{Z}$ is if $m_1^r | d$. The only way this could hold for all r is if $m_1 = \pm 1$, in which case $u \in \mathbb{Z}$.

To prove (b) one proceeds as follows. First, because some of the λ_i's might be equal to each other let us instead write $\mu_1, \dots, \mu_e$ for the list of distinct elements appearing among the λ's, with μ_i appearing r_i times. Our assumption is that for every $k \geq 1$ the number

$$S_k = r_1\mu_1^k + r_2\mu_2^k + \cdots + r_e\mu_e^k$$

lies in $\frac{1}{m}\mathbb{Z}$. Note the matrix equation

$$\begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_e \\ \mu_1^2 & \mu_2^2 & \cdots & \mu_e^2 \\ \vdots & \vdots & \cdots & \vdots \\ \mu_1^e & \mu_2^e & \cdots & \mu_e^e \end{bmatrix} \cdot \begin{bmatrix} r_1\mu_1^k \\ r_2\mu_2^k \\ \vdots \\ r_e\mu_e^k \end{bmatrix} = \begin{bmatrix} S_{k+1} \\ S_{k+2} \\ \vdots \\ S_{k+e} \end{bmatrix}.$$

As the μ_i are distinct, the Vandermonde matrix is invertible; let B denote its inverse. We then obtain

$$r_i\mu_i^k = b_{i1}S_{k+1} + \cdots + b_{ie}S_{k+e} \in \tfrac{1}{m}\mathbb{Z}\langle b_{i1}, \dots, b_{ie}\rangle,$$

where we have used the hypothesis that $S_j \in \frac{1}{m}\mathbb{Z}$ for all j, and where $\mathbb{Z}\langle b_{i1}, \dots\rangle$ denotes the $\mathbb{Z}$-submodule of $\mathbb{C}$ generated by the b_{ij}'s. Therefore

$$\mathbb{Z}[u_i] \subseteq \tfrac{1}{r_i m}\mathbb{Z}\langle b_{i1}, \dots, b_{ie}\rangle,$$

hence $\mathbb{Z}[u_i]$ is finitely-generated as a $\mathbb{Z}$-module; so u_i is integral over $\mathbb{Z}$. $\square$

COROLLARY 2.3. *Let* E *be a field of characteristic zero, let* V *be a finite-dimensional vector space over* E, *and let* $f\colon V \to V$ *be a linear transformation. Suppose that* m *is a positive integer and* $\mathrm{tr}(f^n) \in \frac{1}{m}\mathbb{Z}$ *for every* $n \geq 1$. *Then the* *eigenvalues of f are algebraic integers, and the characteristic polynomial for f lies in $\mathbb{Z}[t]$.*

PROOF. Write $P(t)$ for the characteristic polynomial of f. If the eigenvalues of f are $\lambda_1, \ldots, \lambda_d$, then our assumption is that $\lambda_1^n + \cdots + \lambda_d^n \in \frac{1}{m}\mathbb{Z}$ for every $n \geq 1$. Applying Lemma 2.2(b), we find that the λ_i's are all integral over $\mathbb{Z}$. Since the coefficients of $P(t)$ are integral polynomial expressions in the λ_i's, these coefficients are therefore also integral over $\mathbb{Z}$.

Solving the Newton identities lets us write the coefficients of $P(t)$ as certain polynomial expressions (with rational coefficients) in the numbers $S_n = \lambda_1^n + \cdots + \lambda_d^n$. As these S_n's are rational numbers by assumption, the same will be true for the coefficients of $P(t)$. But then the coefficients of $P(t)$ are both rational and integral over $\mathbb{Z}$, hence they are integers. $\square$

PROOF OF PROPOSITION 2.1. We can write $\pi_i = \frac{z}{m}$ for some integral algebraic cycle z and some integer $m \geq 1$. For convenience write $F = f^*$, and note that $F^n = (f^n)^*$. Then

$$\operatorname{tr}(F^n) = \pm\langle \Gamma_{f^n} \cdot \pi_i \rangle = \pm\tfrac{1}{m} \cdot \langle \Gamma_{f^n} \cdot z \rangle \in \tfrac{1}{m}\mathbb{Z}.$$

Since this holds for every n, Corollary 2.3 says that the coefficients of the characteristic polynomial of F are integers.

???? $\square$

One also has the following simple consequence of $\mathcal{K}\mathrm{u}(X)$:

PROPOSITION 2.4. *Assume that X is a smooth, projective variety over k for which $\mathcal{K}\mathrm{u}(X)$ holds. Then the Betti numbers of X are the same with respect to every Weil cohomology theory.*

PROOF. The ith Betti number is the trace of $id^*\colon H^i(X) \to H^i(X)$, which is the same as $(-1)^i\langle \Delta, \pi_i \rangle$. Since π_i is assumed to be an algebraic cycle, this is an intersection number; and if π_i is independent of the Weil cohomology theory, so is this number. $\square$

The fact that conjecture $\mathcal{K}\mathrm{u}(X)$ has desirable consequences isn't necessarily any reason for believing the conjecture. Is there any reason to believe it? The conjecture is known to be true for projective space, Grassmannians, and other flag manifolds—but for a trivial reason; namely, these varieties have algebraic cell decompositions and as a result *all* of their cohomology groups (and those of their products) are algebraic. That is, for these varieties every element of cohomology is represented by an algebraic cycle.

Perhaps the best evidence for $\mathcal{K}\mathrm{u}(X)$ is the following result of Katz and Messing [**KM**, Theorem 2]:

PROPOSITION 2.5. *Suppose that $H^*(-)$ is a Weil cohomology which has the property that for any smooth, projective scheme X over a finite field $\mathbb{F}_q$, the characteristic polynomial of Frobenius acting on $H^*(X)$ has rational coefficients. Assume as well that the Riemann hypothesis is true for this theory. Then $\mathcal{K}\mathrm{u}(X)$ holds for every such X. In fact, the Künneth components of the diagonal are rational linear combinations of the graphs of powers of the Frobenius map.*

PROOF. Let $P_i(t)$ be the characteristic polynomial of the Frobenius acting on $H^i(X)$. By the Riemann hypothesis, $P_i(t)$ and $P_j(t)$ are relatively prime for $i \neq j$. Let

$$G(t) = \prod_{j \neq i} P_j(t),$$

which is also relatively prime to $P_i(t)$. By the Euclidean algorithm, we can write $1 = A(t)P_i(t) + B(t)G(t)$ for some polynomials $A(t), B(t) \in \mathbb{Q}[t]$. Let $f(t) = 1 - A(t)P_i(t)$, so that $f(t)$ is divisible by $G(t)$ and is congruent to 1 mod $P_i(t)$.

Let Γ be the graph of the Frobenius morphism, regarded as an element of the ring $A^X(X \times X)$. Consider the algebraic cycle $u = f(\Gamma)$. For any j, the map $u^*\colon H^j(X) \to H^j(X)$ is equal to $f(F^*|_{H^j(X)})$. But for $j \neq i$, $P_j(t)$ divides $f(t)$, and by the Cayley-Hamilton theorem $P_j(F^*|_{H^j(X)}) = 0$. So u^* acts as zero on all $H^j(X)$ for $j \neq i$. Likewise, since $f(t)$ is congruent to 1 mod $P_i(t)$ it follows that u^* acts as the identity on $H^i(X)$. In other words, u is exactly the ith Künneth component of the diagonal.

We needed that the coefficients of the $P_i(t)$'s were rational so that the same was true for $f(t)$, which guaranteed that $f(\Gamma)$ was rationally algebraic. $\square$

REMARK 2.6. Because l-adic étale cohomology is a Weil cohomology theory, and because Deligne has proven that the Riemann hypothesis is true, Proposition 2.5 shows that conjecture $\mathcal{K}\mathrm{u}(X)$ *does* hold for all smooth, projective varieties over finite fields.

3. The Lefschetz standard conjecture

Let X be a smooth, projective algebraic variety, and let $\xi \in H^2(X)$ be the class of a hyperplane section. Let $L\colon H^*(X) \to H^{*+2}(X)$ be the map $a \mapsto a \cdot \xi$. Using the Hard Lefschetz Theorem, we obtain a primitive decomposition for $H^*(X)$ in the usual way: for $0 \leq i \leq X$, define

$$PH^i(X) = \ker\big[L^{X-i+1}\colon H^i(X) \to H^{2X-2i+2}(X)\big].$$

Then for all $0 \leq i \leq 2X$,

$$H^i(X) = PH^i(X) \oplus L[PH^{i-2}(X)] \oplus \cdots$$

In other words, for any $a \in H^i(X)$ there is a unique representation of a in the form

$$a = a_0 + La_1 + L^2a_2 + \cdots \tag{3.1}$$

where each a_j lies in $PH^{i-2j}(X)$.

Let us now consider the following operators:

(i) $\pi_i\colon H^*(X) \to H^*(X)$, projection onto $H^i(X)$. That is, if $a \in H^j(X)$ then $\pi_i(a) = \delta_{i,j}a$.

(ii) $\Lambda\colon H^*(X) \to H^*(X)$, defined by

$$\Lambda(a) = a_1 + La_2 + \cdots$$

(iii) $p_j\colon H^*(X) \to H^*(X)$, defined to be nonzero only on $H^j(X)$ and to satisfy

$$p_j(a) = \begin{cases} a_0 & \text{if } 0 \leq j \leq X, \\ a_{2X-j} & \text{if } X < j \leq 2X. \end{cases}$$

More intuitively, when $0 \leq j \leq X$ the operator p_j is simply projection onto $PH^j(X)$. When $X < j \leq 2X$ there is no primitive component of $H^j(X)$, and the operator p_j instead projects onto the primitive component of $H^{2X-j}(X)$. In all cases $p_j(a)$ is obtained by taking the component having the least number of L's in the primitive decomposition of a, and then removing all of those L's.

(iv) $*\colon H^*(X) \to H^*(X)$, defined as follows. If $b = L^k b_0$ for $b_0 \in PH^{i-2k}(X)$, then

$$*(b) = (-1)^{\binom{i+1}{2}} \cdot (-1)^k \cdot L^{X-i+k} b_0 = (-1)^{\binom{i-2k+1}{2}} \cdot L^{X-i+k} b_0.$$

Note that if $0 \le i \le X$ and $a \in H^i(X)$ has primitive decomposition as in (3.1), then

$$*(x) = (-1)^{\binom{i+1}{2}} \cdot L^{X-i}[a_0 - La_1 + L^2 a_2 - \cdots].$$

In other words, $*\colon H^i(X) \to H^{2X-i}(X)$ is a "twisted" form of L^{X-i} in which one alternates the signs on the different pieces of the primitive decomposition.

The operators L, Λ, π_i, p_j, and $*$ will be called the **standard operators of Hodge theory**. Note that each will be represented by a class in $H^*(X \times X)$.

We define the "Lefschetz standard conjecture" to be the following:

$\Lambda(X,\xi)$: The standard operators of Hodge theory are all represented by algebraic cycles in $H^*(X \times X)$.

We will see below that this conjecture has interesting and useful consequences. Note that it contains the Künneth conjecture, but seems to go much further. Our phrasing of the conjecture is perhaps a bit too broad; with a little work we can simplify it to make it seem more approachable:

Lemma 3.2. *The operator L is algebraic; indeed, it is represented by the algebraic cycle $\Delta_!(\xi_X) \in H^{2X+2}(X \times X)$, where $\Delta\colon X \to X \times X$ is the diagonal.*

Proof. We must prove that $[\Delta_!(\xi)]^*(z) = \xi \cdot z$, for all $z \in H^*(X)$. But

$$\begin{aligned}[\Delta_!(\xi)]^*(z) = z \circ \Delta_!(\xi) = (\pi_2)_![(z \otimes 1) \cdot \Delta_!(\xi)] &= (\pi_2)_!\Delta_![\Delta^*(z \otimes 1) \cdot \xi] \\ &= z \cdot \xi \\ &= \xi \cdot z,\end{aligned}$$

where in the second-to-last equality we have used $\pi_2 \circ \Delta = id$ and $\Delta^*(z \otimes 1) = z$. □

Proposition 3.3. *For any smooth, projective variety X the following are equivalent:*

(a) The operator Λ is algebraic;
(b) The operator $*$ is algebraic;
(c) The Lefschetz standard conjecture holds for X.

In other words, if either Λ or $*$ is algebraic then so are all the other standard operators of Hodge theory.

Proof. Consider the following $\mathbb{Q}$-subalgebras of $\mathrm{End}(H^*(X))$:

$$\mathbb{Q}\langle L, \Lambda\rangle, \quad \mathbb{Q}\langle L, *\rangle, \quad \mathbb{Q}\langle L, p_X, p_{X+1}, \ldots, p_{2X}\rangle.$$

We will prove that these subalgebras are equal, and that they all contain π_i for $0 \le i \le 2X$ and p_j for $0 \le j \le X$. Therefore they contain all the standard operators of Hodge theory. Since we know by Lemma 3.2 that L is algebraic, it follows that if Λ (or $*$) is algebraic, so are all the other standard operators.

Let B denote $\mathbb{Q}\langle L, p_X, p_{X+1}, \ldots, p_{2X}\rangle$. One readily checks the formulas

$$p_0 = p_X L^X, \qquad p_1 = p_{X-1}L^{X-1}$$

and in general $p_i = p_{X-i}L^{X-i}$ when $0 \leq i < X$. These formulas show that $p_0, p_1, \ldots, p_{X-1} \in B$.

Next verify that $L^a p_{2x-b} L^{X-a-b}$ is the projection from $H^*(X)$ onto $L^a PH^b(X)$. From this one obtains the formulas

$$\pi_0 = p_0, \qquad \pi_1 = p_1, \qquad \pi_2 = p_2 + Lp_{2x}L^{X-1}, \qquad \pi_3 = p_3 + Lp_{2X-1}L^{X-2}$$
$$\pi_4 = p_4 + Lp_{(2X-2)}L^{X-1} + L^2 p_{(2X)}L^{X-2}$$

and in general

$$\pi_i = p_i + Lp_{2X-i+2}L^{X-i+1} + L^2 p_{2X-i+4}L^{X-i+2} + \cdots$$

for $0 \leq i \leq X$. Likewise, when $X < i \leq 2X$ we can write

$$\pi_i = L^{i-X}p_i + L^{i-X+1}p_{i+2}L + \cdots$$

These identities show that each π_i belongs to B.

Similar considerations show that $* \in B$. More precisely, when $0 \leq i \leq X$ we can write

$$* \circ \pi_i = \pm L^{X-i}p_i \pm L^{X-i}(Lp_{2X-i+2}L^{X-i+1}) \pm \cdots$$

where the signs can be determined but are of no consequence to us. A similar formula holds in the case $X < i \leq 2X$, and so we have that $* \circ \pi_i \in B$ for all i. Since $* = * \circ 1 = * \circ (\pi_0 + \pi_1 + \cdots + \pi_{2X})$ and each $* \circ \pi_i$ belongs to B, it follows that $*$ belongs to B.

The last paragraph proved that $* \in B$. The fact that $\Lambda = *L*$ now gives

$$\mathbb{Q}\langle L, \Lambda\rangle \subseteq \mathbb{Q}\langle L, *\rangle \subseteq B.$$

To complete our proof it suffices to show that $Q\langle L, \Lambda\rangle$ contains $p_X, p_{X+1}, \ldots, p_{2X}$. Note that $p_{2X} = \Lambda^X$ and $p_{2X-1} = \Lambda^{X-1} - L\Lambda^X - \Lambda^X L$. One can find identities similar to these for each p_{2X-j}, but it becomes unpleasant to do this by brute force. Instead recall that $L^a p_{2X-b}L^{X-a-b}$ is the projection onto $L^a PH^b(X)$. So $(id - L^a p_{2X-b}L^{X-a-b})$ is the identity on all factors of the primitive decomposition except $L^a PH^b(X)$, on which it vanishes. From this it is easy to check that for $0 \leq j \leq X$ one has

$$p_{2X-j} = \Lambda^{X-j} \circ \prod_{a,b} \left(id - L^a p_{2x-b}L^{X-a-b}\right),$$

where in the product we require $a \geq X - j$ and $b < j$ (and the product is the composition product). It now follows by a reverse induction that $p_X, p_{X+1}, \ldots, p_{2X}$ all lie in $\mathbb{Q}\langle L, \Lambda\rangle$, and this completes our proof. □

We also note the following consequence of the Lefschetz standard conjecture.

PROPOSITION 3.4. *Assume $\Lambda(X, \xi)$. Then for each $0 \leq i \leq X$, the map*

$$L^{X-2i} \colon A^i(X) \to A^{X-i}(X)$$

is an isomorphism.

Proof. One simply considers the square

$$\begin{tikzcd} H^{2i}(X) \arrow[r, "L^{X-2i}"] & H^{2X-2i}(X) \\ A^i(X) \arrow[u, hook] \arrow[r] & A^{X-i}(X), \arrow[u, hook] \end{tikzcd}$$

which exists because L takes algebraic classes to algebraic classes. The upper horizontal map is an isomorphism because $H^*(-)$ is assumed to satisfy the Hard Lefschetz Theorem. Therefore the lower horizontal map is injective.

The map Λ^{X-2i} is the inverse of L^{X-2i}, and the conjecture $\Lambda(X,\xi)$ says that this operator is algebraic. Therefore Λ^{X-2i} also takes algebraic classes to algebraic classes, which proves that $A^i(X) \to A^{X-i}(X)$ is surjective. $\square$

4. Algebraic preliminaries

Before proceeding further with our treatment of the standard conjectures, we pause briefly in this section to develop some purely algebraic results.

Let E be a field, V be a finite-dimensional vector space over E, and $h\colon V \to V$ a linear transformation. Consider the algebra $E[h] \subseteq \mathrm{End}(V)$. The map h is said to be **semisimple** if the algebra $E[h]$ is semisimple; since $E[h]$ is commutative, this latter condition is equivalent to saying that $E[h]$ is a product of fields.

It turns out that the map h is semisimple precisely when h is diagonalizable over the algebraic closure $\bar{E}$ of E. This is very classical, but we will recall the argument here since it is brief. Write $E[h] \cong E[x]/(p(x))$ where $p(x)$ is monic, and recall that $p(x)$ is called the minimum polynomial of h. If $P(x)$ denotes the characteristic polynomial of h, then by Cayley-Hamilton we know $P(h) = 0$ and therefore $p(x)$ divides $P(x)$. So $p(x) = (x-\lambda_1)^{e_1}\cdots(x-\lambda_n)^{e_n}$ where the λ_i's are the eigenvalues of h and each e_i is a positive integer less than or equal to the multiplicity of λ_i as an eigenvalue. In fact it's easy to see that the e_i's are the maximal sizes of the Jordan λ_i-blocks in the Jordan canonical form for h; so h is diagonalizable over $\bar{E}$ precisely when all the e_i's are equal to 1. But note that

$$\begin{aligned} E[h]\otimes_E \bar{E} \cong \bar{E}[x]/(p(x)) &\cong \bar{E}[x]/(x-\lambda_1)^{e_1}\times\cdots\times\bar{E}[x]/(x-\lambda_n)^{e_n} \\ &\cong \bar{E}[t]/(t^{e_1})\times\cdots\times\bar{E}[t]/(t^{e_n}). \end{aligned}$$

Clearly this is semisimple only when $e_1 = e_2 = \cdots = e_n = 1$. To complete the argument, just observe that a commutative E-algebra (like $E[h]$) is semisimple if and only if it becomes semisimple after being tensored with $\bar{E}$ (or any field extension or E, for that matter).

Note that another phrasing of what we just showed is that a linear transformation is semisimple precisely when its minimum polynomial has no linear factors of multiplicity greater than 1 (over an algebraically closed extension field).

The space $E[g] \cong E[x]/(p(x))$ is finite-dimensional, and multiplication by g gives an endomorphism $L_g\colon E[g] \to E[g]$. The properties of L_g are closely related to the properties of the original map g:

Proposition 4.1. *Let $g\colon V \to V$ be an endomorphism of a finite-dimensional vector space over E. Let $L_g\colon E[g] \to E[g]$ be the map given by $h \mapsto gh$. Then the eigenvalues of g (over an algebraically closed extension field) are the same as the eigenvalues of L_g, and g is semisimple if and only if L_g is semisimple.*

PROOF. We replace L_g by the isomorphic map $L_x\colon E[x]/(p(x)) \to E[x]/(p(x))$, where $p(x)$ is the minimum polynomial of g. It is immediate that the minimum polynomial of L_x is $p(x)$. Since L_g and g therefore have the same minimum polynomial, the statements in the proposition follow immediately. $\square$

Finally, we recall the following classical result:

PROPOSITION 4.2. *Suppose that $E \subseteq \mathbb{R}$ is a subfield, and $g\colon V \to V$ is an endomorphism of a finite-dimensional vector space over E. Assume there is a symmetric, bilinear form $(-,-)$ on V which is positive-definite and which is preserved by g: that is, where $(gx, gy) = (x, y)$ for all $x, y \in V$. Then g is semisimple and its eigenvalues (over $\mathbb{C}$) are all of norm 1.*

PROOF. First, we may assume $E = \mathbb{R}$ by extending scalars. Then since the form is positive-definite, there is a basis for V with respect to which the form is the usual norm form on $\mathbb{R}^n$. So we might as well assume $V = \mathbb{R}^n$ and the form is standard inner product.

Now tensor with the complex numbers, to obtain $g\colon \mathbb{C}^n \to \mathbb{C}^n$ which is unitary. The statement about the eigenvalues is now evident: if $g(x) = \lambda x$ where $x \in \mathbb{C}^n$ is nonzero then

$$(x, x) = (gx, gx) = (\lambda x, \lambda x) = \lambda\bar{\lambda}(x, x).$$

Since $(x, x) \neq 0$ we have $|\lambda| = 1$.

To prove that g is diagonalizable, choose an eigenvector x_1 of $\mathbb{C}^n$. Let $V_1 = \langle x_1\rangle^{\perp}$. Since g is unitary, g restricts to a map $V_1 \to V_1$. Now pick an eigenvector $x_2 \in V_1$ for g,and let $V_2 = \langle x_1, x_2\rangle^{\perp} \subseteq \mathbb{C}^n$. Continuing in this way, one produces a basis for $\mathbb{C}^n$ which diagonalizes g. $\square$

The next result is of a slightly different nature. We include it here because it will be useful in the next section.

PROPOSITION 4.3. *Let E be a field, and let V be a finite-dimensional vector space over E with a nondegenerate symmetric bilinear form $(-,-)$. Given $f\colon V \to V$, there is a unique map $f^\dagger\colon V \to V$ with the property that*

$$(f^\dagger(a), b) = (a, f(b))$$

for all $a, b \in V$. The characteristic polynomials of $f^\dagger$ and f are identical.

PROOF. The bilinear form on V gives an isomorphism $\phi\colon V \to V^*$ by sending a to the functional $(a, -)$. One checks readily that $f^\dagger$ is equal to the composite $\phi^{-1} \circ f^* \circ \phi$; that is, there is a commutative diagram

$$\begin{tikzcd} V \arrow[r] \arrow[d, "f^\dagger"'] & V^* \arrow[d, "f^*"] \\ V \arrow[r] & V^*. \end{tikzcd}$$

The above square immediately implies that the characteristic polynomials for $f^\dagger$ and f^* are identical, and the latter is of course the same as the characteristic polynomial for f (the polynomials for a matrix A and its transpose being equal). $\square$

5. The Hodge standard conjecture

Let X be a smooth, projective variety over a field k, and let $\xi \in CH^1(X)$ be the class of a hyperplane section. We will also write ξ for the corresponding element of $A^2(X) \subseteq H^2(X)$. Let $L\colon H^i(X) \to H^{i+2}(X)$ be the Lefschetz operator $x \mapsto x \cdot \xi$, and note that when i is even this restricts to an operator $A^i(X) \to A^{i+2}(X)$. Define the **primitive component** of $A^i(X)$ to be

$$PA^i(X) = \{x \in A^i(X) \,|\, L^{X-i+i}(x) = 0\}.$$

In analogy with our experience in topology (see Chapter 2), one might make the following conjecture:

IH(X, ξ) : For each even number $0 \leq i \leq X$, the pairing on $PA^i(X)$ given by $a, b \mapsto (-1)^i \langle L^{X-i}a, b\rangle$ is positive-definite.

We will call this the **Hodge standard conjecture** for X, or the **Hodge index conjecture**. The acronym IH is supposed to represent 'Hodge' and 'Index' (writing HI(X, ξ) looks a little too much like a homology group!) Note that the conjecture depends on the class ξ, although we will sometimes tend to suppress this in our discussion.

5.1. An involution on $H^*(X \times X)$**.** For $u \in H^*(Y \times X)$, we can define another class $\bar{u} \in H^*(X \times Y)$ by the formula

$$\left\langle u^*(a), b\right\rangle_{H,X} = \left\langle a, \bar{u}^*(b)\right\rangle_{H,Y}$$

(here $a \in H^*(Y)$ and $b \in H^*(X)$). As the Hodge pairing is nondegenerate, this uniquely determines $\bar{u}$.

Lemma 5.2. *If u is even-dimensional then $\bar{u}^* = *_X \circ (u^t)^* \circ *_Y$.*

Proof. One simply computes:

$$\begin{aligned}\Big\langle a, (*_X \circ u^t \circ *_Y)(b)\Big\rangle_{H,Y} = \Big\langle a, *_Y(u^t(*_X(b)))\Big\rangle_{H,Y} &= \left\langle a, u^t(*_X(b))\right\rangle_Y \\ &= \left\langle u^*(a), *_X(b)\right\rangle_X \\ &= \langle u^*(a), b\rangle_{H,X}.\end{aligned}$$

□

Lemma 5.3. *If $u \in H^*(Y \times X)$ and $v \in H^*(Z \times Y)$ then $\overline{v \circ u} = \bar{u} \circ \bar{v}$.*

Proof. This is an easy exercise using adjointness and $(v \circ u)^* = u^* \circ v^*$. □

If V is a graded vector space and $h\colon V \to V$ is a degree zero linear map, there are two reasonable definitions for $\mathrm{Tr}(h)$. One is the usual trace, where one ignores the grading. The other is the "graded trace", where the traces on the odd-dimensional pieces of V are counted with a negative sign. Both definitions have their uses. For us, we will always use the former definition: so $\mathrm{Tr}(h)$ is the classical trace, which is also the sum (without negative signs) of the classical traces on each homogeneous component of V.

Define a bilinear form on $H^*(Y \times X)$ by

$$(u, v) = \mathrm{Tr}(\bar{u} \circ v).$$

This is clearly bilinear, and it is also symmetric:

$$(v,u) = \mathrm{Tr}(\bar{v}\circ u) = \mathrm{Tr}\Big(\overline{(\bar{v}\circ u)}\Big) = \mathrm{Tr}(\bar{u}\circ v) = (u,v),$$

where in the second equality we have used Proposition 4.3.

THEOREM 5.4. *Let X and Y be smooth, projective algebraic varieties with hyperplane sections $\xi_X \in H^2(X)$ and $\xi_Y \in H^2(Y)$. Assume the following conjectures:*
(i) $\Lambda(X,\xi_X)$ and $\Lambda(Y,\xi_Y)$,
(ii) $IH(Y\times X, \xi_Y\otimes 1 + 1\otimes \xi_X)$.
Then for every nonzero $u \in H^*(Y\times X)$ which is algebraic, $\mathrm{Tr}(\bar{u}\circ u)$ is a positive rational number. Consequently, the form $(u,v) = \mathrm{Tr}(\bar{u}\circ v)$ on the rational vector space $A^*(Y\times X)$ is positive-definite.

Before proving this let us establish a helpful lemma. Let X and Y be smooth, projective algebraic varieties with hyperplane sections $\xi_X \in H^2(X)$ and $\xi_Y \in H^2(Y)$. Then

$$\xi_{X\times Y} = \xi_X\otimes 1 + 1\otimes \xi_Y$$

is a hyperplane section for $X\times Y$.

LEMMA 5.5. *If $u \in PH^i(Y)\otimes PH^j(X)$ then*

$$*_{Y\times X}(u) = (-1)^{ij}\tbinom{X+Y-i-j}{Y-i}\cdot[(*_Y\otimes *_X)(u)].$$

PROOF. By the definition of the Hodge $*$-operator,

$$\begin{aligned} *_{Y\times X}(u) &= (-1)^{\binom{u+1}{2}}\cdot(\xi_{Y\times X})^{Y+X-u}\cdot u\\ &= (-1)^{\binom{u+1}{2}}\cdot\sum_k \tbinom{X+Y-u}{k}(\xi_Y^k\otimes\xi_X^{X+Y-u-k})u. \end{aligned}$$

But since $u \in PH^i(Y)\otimes PH^j(X)$, the terms inside the sum vanish unless $k \le Y-i$ and $X+Y-u-k \le X-j$. The second equality may be rewritten as $k \ge Y-u+j = Y-i$ (using that $i+j=u$). So there is only one non-vanishing term in the sum, namely where $k = Y-i$. We therefore have

$$\begin{aligned} *_{Y\times X}(u) &= (-1)^{\binom{u+1}{2}}\cdot\tbinom{X+Y-u}{Y-i}\cdot(\xi_Y^{Y-i}\otimes\xi_X^{X-u+i})\cdot u\\ &= (-1)^{\binom{u+1}{2}+\binom{i+1}{2}+\binom{j+1}{2}}\cdot\tbinom{X+Y-u}{Y-i}\cdot(*_Y\otimes *_X)(u). \end{aligned}$$

Using that $|u| = i+j$, the only thing left is to see that

$$\tbinom{i+j+1}{2}+\tbinom{i+1}{2}+\tbinom{j+1}{2} \equiv ij \bmod 2.$$

We leave this to the reader. $\square$

Let $q_X^{a,b}$ be the projection $H^*(X)\to L^aPH^b(X)$ which is zero on all pieces of the Lefschetz decomposition except for $L^aPH^b(X)$.

LEMMA 5.6. *The following are true:*

$$\overline{\Lambda_X} = L_X, \qquad \overline{q_X^{a,b}} = q_X^{a,b}, \qquad \textit{and} \qquad (*_X)^t = *_X.$$

PROOF. For the first, compute that

$$\langle a,\overline{\Lambda}(b)\rangle_H = \langle\Lambda(a),b\rangle_H = \langle *\Lambda(a),b\rangle = \langle L(*a),b\rangle = \langle *a,L(b)\rangle = \langle a,L(b)\rangle_H.$$

Here we have used that $\Lambda = *L*$ and $L(x)\cdot y = \xi xy = x\xi y = x\cdot L(y)$.

For the second, recall that the different summands of $H^b(X)$ in the Lefschetz decomposition are orthogonal to each other with respect to $\langle -,-\rangle_H$. So

$$\langle q^{a,b}(z), w\rangle_H = \langle q^{a,b}(z), q^{a,b}(w)\rangle_H = \langle z, q^{a,b}(w)\rangle_H.$$

This exactly proves that $\overline{q^{a,b}} = q^{a,b}$.

For the final identity, one has $\langle a, (*_X)^t(b)\rangle = \langle *(a), b\rangle = \langle a, *b\rangle$ where in the first equality we have used Exercise 5.20 together with the fact that $*_X$ is even-dimensional in $H^*(X\times X)$. $\square$

Proof of Theorem 5.4. First assume that $u \in PH^i(Y)\otimes PH^j(X)$, for some i and j. Note that $i+i$ will be even, since u is algebraic. We have

$$\begin{aligned}(\bar{u})^t = (*_X \circ u^t \circ *_Y)^t = (*_Y)^t \circ u \circ (*_X)^t = *_Y \circ u \circ *_X &= (*_Y \otimes *_X)(u)\\ &= \frac{(-1)^{ij}}{B}\cdot \Big[*_{Y\times X}(u)\Big]\end{aligned}$$

where $B = \binom{X+Y-i-j}{X-i}$.

Now, $(\bar{u}\circ u)^*$ is a map $H^j(X)\to H^j(X)$ (that is, it is zero on all $H^k(X)$ for $k\neq j$). So by the Lefschetz Trace Formula,

$$\begin{aligned}\mathrm{Tr}(\bar{u}\circ u) = (-1)^j \left\langle (\bar{u})^t, u\right\rangle_{Y\times X} &= \frac{(-1)^{ij+j}}{B}\left\langle (*_{Y\times X})(u), u\right\rangle_{Y\times X}\\ &= \frac{(-1)^{ij+j}}{B}\left\langle u,u\right\rangle_{H,Y\times X}.\end{aligned}$$

But $IH(Y\times X)$ implies that $\langle -,-\rangle_H$ is positive definite, so $\langle u,u\rangle_H > 0$. Finally, $j\equiv j^2 \pmod 2$, so $(-1)^{ij+j} = (-1)^{ij+j^2} = (-1)^{j(i+j)} = 1$ because $i+j$ is even. So $\mathrm{Tr}(\bar{u}\circ u) > 0$, and the proof for this case is complete.

We tackle the general case of $u\in H^*(Y\times X)$ by reducing it to the case handled above. First, we may obviously assume that u is homogeneous. Let $q_X^{a,b}\colon H^*(X)\to H^*(X)$ be the projection of the Lefschetz decomposition onto $L^aPH^b(X)$. Then we may write

$$u = \sum_{a,b,c,d} q_Y^{c,d}\circ u \circ q_X^{a,b}.$$

We then have

$$\bar{u} = \sum_{a,b,c,d} q_X^{a,b}\circ \bar{u}\circ q_Y^{c,d}$$

and

$$\bar{u}\circ u = \sum_{a,b,c,d,a',b',c',d'} q_X^{a,b}\circ \bar{u}\circ q_Y^{c,d}\circ q_Y^{c',d'}\circ u\circ q_X^{a',b'}.$$

The composite of the two q operators in the middle is zero unless $c=c'$ and $d=d'$, and the entire composite has zero trace unless $a=a'$ and $b=b'$. So

$$\mathrm{Tr}(\bar{u}\circ u) = \sum_{a,b,c,d}\mathrm{Tr}(\overline{u_{a,b,c,d}}\circ u_{a,b,c,d})$$

where $u_{a,b,c,d} = (q_Y^{c,d}\circ u\circ q_X^{a,b})$. By conjectures $\Lambda(X)$ and $\Lambda(Y)$, all of the q operators are algebraic; therefore each $u_{a,b,c,d}$ is algebraic. Hence, it suffices to prove the theorem for all of the $u_{a,b,c,d}$ classes. That is, we may assume that u itself is of this form, i.e. that u^* factors as

$$H^*(Y)\xrightarrow{q} L^cPH^d(Y)\longrightarrow L^aPH^b(X)\hookrightarrow H^b(X).$$

For the final reduction, let $v = \Lambda_Y^{Y-d-c} \circ u \circ \Lambda_X^a$. One should think of this in terms of the following picture:

$$\begin{tikzcd} L^{Y-d}PH^d(Y) \arrow[r, shift left, "\Lambda_Y^{Y-d-c}"] & L^cPH^d(Y) \arrow[l, shift left, "L^{Y-d-c}"] \arrow[d, "u^*"] & PH^d(Y) \arrow[l, "L^c"'] \\ & L^aPH^b(X) \arrow[r, "\Lambda_X^a"] & PH^b(X). \end{tikzcd}$$

So v^* is a map $L^{Y-d}PH^d(Y) \to PH^b(X)$. Its codomain is as primitive as possible, and its domain is as "non-primitive" as possible (that is, the domain is Hodge dual to the primitives). We calculate that

$$\bar{v} = (\overline{\Lambda_X})^a \circ \bar{u} \circ (\overline{\Lambda_Y})^{Y-d-c} = L_X^a \circ \bar{u} \circ L_Y^{Y-d-c}$$

and so

$$\bar{v} \circ v = L_X^a \circ \bar{u} \circ L_Y^{Y-d-c} \circ \Lambda_Y^{Y-d-c} \circ u \circ \Lambda_X^a = L_X^a \circ (\bar{u} \circ u) \circ \Lambda_X^a.$$

Therefore

$$\mathrm{Tr}(\bar{v} \circ v) = \mathrm{Tr}(\bar{u} \circ u).$$

Again, the conjectures $\Lambda(X)$ and $\Lambda(Y)$ imply that the operators L and Λ are algebraic, hence v is algebraic. We claim that $v \in PH^{Y-d}(Y) \otimes PH^b(X)$, and this will complete the proof because this case has already been handled.

To justify the claim about v, let $\beta_1, \dots, \beta_k$ be a basis for $PH^b(X)$ and extend this to a basis $\beta_1, \dots, \beta_r$ for $H^b(X)$. Likewise, let $\gamma_1, \dots, \gamma_l$ be a basis for $L^{Y-d}PH^{2Y-d}(Y)$ and extend it to a basis $\gamma_1, \dots, \gamma_s$ of $H^{2Y-d}(Y)$. Finally, let $\{\hat{\gamma}_p\}$ be the dual basis for $H^d(Y)$, defined by $\langle \hat{\gamma}_p, \gamma_q \rangle = \delta_{p,q}$. Note that $\hat{\gamma}_1, \dots, \hat{\gamma}_l \in PH^d(Y)$.

As v^* is a map $H^{2Y-d}(Y) \to H^b(X)$, write $v^*(\gamma_q) = \sum_i c_{qi}\beta_i$ for $c_{q,i} \in \mathbb{Q}$. Then by Lemma 5.14 one has

$$v = (-1)^d \sum_{q,i} c_{q,i}(\hat{\gamma}_q \otimes \beta_i).$$

Our construction of v^* gives that $v^*(\gamma_q) = 0$ for $q > l$ and that the image of v^* is contained in $PH^b(X)$. So $c_{q,i} = 0$ for $i > k$ or $q > l$, which gives immediately that $v \in PH^d(Y) \otimes PH^b(X)$. □

5.7. The standard conjectures and their consequences. By the "Standard Conjectures" we mean:

- The Künneth conjecture $\mathcal{K}\mathrm{u}(X)$;
- The Lefschetz conjecture $\Lambda(X)$;
- The Hodge standard conjecture $\mathrm{IH}(X, \xi)$.

Of course we have seen that the first of these is a consequence of the second, and so technically doesn't have to be listed separately.

Proposition 5.8. *Let X be a smooth, projective algebraic variety over k of even dimension. Let $u \in H^X(X \times X)$ be an algebraic cycle such that $u^*(\xi) = q\xi$ for some hyperplane section $\xi \in H^2(X)$ and some positive rational number q. Assuming the Standard Conjectures hold, then for each i the induced map $u^* \colon H^i(X) \to H^i(X)$ is semisimple and the eigenvalues are algebraic numbers which have absolute norm $q^{i/2}$.*

PROOF. First consider the algebra $A = A^X(X \times X) \subseteq H^X(X \times X)$ consisting of the algebraic degree 0 correspondences. This is a finite-dimensional $\mathbb{Q}$-algebra. Define a symmetric bilinear form on this algebra by

$$(a, b) = \operatorname{Tr}(\bar{a}b).$$

We know by Proposition 5.4 that this form takes values in $\mathbb{Q}$ and is positive-definite.

Let $g\colon H^*(X) \to H^*(X)$ be the map which on $H^i(X)$ sends

$$a \mapsto \frac{u^*(a)}{q^{i/2}}.$$

This only makes sense if $q^{1/2}$ belongs to the coefficient field E, but if necessary we can extend E so that this is true. Note that g is a ring map, and that $g(\xi) = \xi$. Let $F = \mathbb{Q}(q^{1/2})$.

Write $A_F = A \otimes_{\mathbb{Q}} F \subseteq H^X(X \times X)$, and note that the algebraic cycle representing g lies in A_F. This is true because ????

Our goal is to show that g is semisimple, and that its eigenvalues are algebraic integers having absolute norm 1. By Proposition 4.1 it will be enough to prove the same for the action of g on the subalgebra $F[g] \subseteq A_F$.

We claim that $\bar{g}g = 1$. Granting this for the moment, it implies that

$$(ga, gb) = \operatorname{Tr}(\overline{ga} \cdot gb) = \operatorname{Tr}(\bar{a}\bar{g} \cdot gb) = \operatorname{Tr}(\bar{a}b) = (a, b).$$

So g preserves the positive-definite form $(-, -)$ on $F[g]$, and therefore by Proposition 4.2 g is semisimple and its eigenvalues have absolute norm 1.

Our final task is to verify that $\bar{g}g = 1$. Let $d = \dim X$. Since $g(\xi) = \xi$ and g is a ring map, $g^*(\xi^d) = \xi^d$. So the map $g^*\colon H^{2d}(X) \to H^{2d}(X)$ is the identity. If $a \in H^i(X)$, then by duality there is a $b \in H^{2d-i}(X)$ such that $ab \neq 0$. Then $g^*(a)g^*(b) = g^*(ab) = ab \neq 0$, so $g^*(a) \neq 0$. Hence g^* is injective, and since $H^*(X)$ is finite-dimensional it follows that g^* is an automorphism.

For any $a \in H^i(X)$ and $b \in H^{2d-i}(X)$,

$$\begin{aligned} \langle (g^*)^{-1}a, b\rangle = \eta_X\big((g^*)^{-1}(a) \cdot b\big) &= \eta_X\Big(g^*\big((g^*)^{-1}(a) \cdot b\big)\Big) \\ &= \eta_X(a \cdot g^*(b)) \\ &= \langle a, g^*(b)\rangle. \end{aligned}$$

This shows that $g^{-1} = g^t$ in A.

Finally, since $g^*(\xi) = \xi$ it follows that g^* preserves the primitive decomposition of $H^*(X)$. Therefore $(g^*)^{-1}$ likewise preserves the decomposition, so $(g^*)^{-1}$ commutes with the Hodge $*$ operator:

$$*_X \circ (g^*)^{-1} \circ *_X = (g^*)^{-1}.$$

Note that we then have the same identity involving g^t. So we finally compute that

$$\bar{g} = *_X \circ g^t \circ *_X = g^t = g^{-1}$$

and we are done. □

COROLLARY 5.9. *Assume the Standard Conjectures hold for smooth, projective varieties over $\mathbb{F}_q$. Then the Riemann Hypothesis also holds for such varieties.*

PROOF. If a variety X gave a counterexample to the Riemann hypothesis, then $X \times \mathbb{P}^1$ would also be a counterexample (using the Künneth Theorem). So it is enough to prove that the Riemann hypothesis holds for all even-dimensional varieties.

If X is smooth, projective, and even-dimensional, then the graph of Frobenius gives an algebraic cycle $\Gamma_f \in H^{2X}(X \times X)$. Moreover, we know that $\Gamma_f^*(\xi) = q\xi$ for any hyperplane section ξ. Proposition 5.8 immediately gives the desired result. $\square$

We also record the following interesting consequence, for later use.

PROPOSITION 5.10. *Let X be a smooth, projective algebraic variety over k and assume the Standard Conjectures hold. Let B be any subalgebra of $A^X(X \times X)$ which is closed under the operation $u \mapsto \bar{u}$. Then B is a semisimple algebra. In particular, this holds when $B = A^X(X \times X)$.*

PROOF. To say that B is semisimple is to say that its Jacobson radical is zero. So let u belong to the Jacobson radical of B, and assume $u \neq 0$. Since $\bar{u} \in B$, the element $\bar{u} \circ u$ also belongs to the Jacobson radical. But B is finite-dimensional over $\mathbb{Q}$, therefore artinian, and so the Jacobson radical is nilpotent. In particular, $\bar{u} \circ u$ is nilpotent. Note that $\mathrm{Tr}(\bar{u} \circ u) > 0$, and so $\bar{u} \circ u \neq 0$. Choose the smallest m such that $(\bar{u} \circ u)^{2^m} = 0$ and let $v = (\bar{u} \circ u)^{2^{m-1}}$. Then v is nonzero and

$$\bar{v} \circ v = (\bar{u} \circ u)^{2^m} = 0,$$

and so $\mathrm{Tr}(\bar{v} \circ v) = 0$. This is a contradiction, because v is rationally algebraic. $\square$

6. Hodge decompositions in characteristic p

???? In our discussion of the Weil conjectures from Chapter 1, we progressed from trying to explain formulas for counting points to speculating about the existence of a cohomology theory for algebraic varieties. As part of this business, it has been natural to ask ourselves what properties of singular cohomology could be expected to hold for our algebraic cohomology theory. At this point we have seen that Poincaré duality is a reasonable expectation, and that Lefschetz theorems and the Hodge Index Theorem would be highly desirable properties. We have also seen that the geometry of characteristic p varieties forces certain differences between our algebraic theory and the singular theory: namely, we saw that the coefficient field for our algebraic theory could not be any subfield of $\mathbb{R}$.

Now we would like to turn the discussion to the question of Hodge decompositions. Is it reasonable to expect that our sought-after algebraic cohomology groups would admit some kind of Hodge decompositions? It seems to be Weil who first observed that this *cannot* happen—at least not in the form one would expect—again because of some peculiarities in characteristic p geometry. We will now try to explain this. (For the attribution to Weil, see [**Kl1**, p. 360]).

Let C be a supersingular elliptic curve over an algebraically closed field k of characteristic p. We know that our expected algebraic cohomology groups must be $H^0(C) = H^2(C) = E$ and $H^1(C) = E^2$, where E is the coefficient field of the theory. Consider the algebraic surface $S = C \times C$. By the Künneth theorem we would then have:

i	0	1	2	3	4
$H^i(S)$	E	E^4	E^6	E^4	E

If we speculate about possible Hodge decompositions, it seems clear that for $H^*(C)$ the decomposition would need to be:

$$\begin{array}{llll} H^2: & & E & \\ H^1: & E & & E \\ H^0: & & E. & \end{array}$$

That is, we would need to have $H^{0,1}(C) = H^{1,0}(C) = E$, and likewise $H^{0,0}(C) = H^{1,1}(C) = E$. By Künneth this would then give the following decomposition for $H^*(S)$:

$$\begin{array}{llllll} H^4: & & & E & & \\ H^3: & & E^2 & & E^2 & \\ H^2: & E & & E^4 & & E \\ H^1: & & E^2 & & E^2 & \\ H^0: & & & E. & & \end{array}$$

In particular, note that $H^{1,1}(S) = E^4$. This is what will lead to our contradiction.

Recall the cycle class map $Z^i(S) \to H^{2i}(S)$. In analogy with classical Hodge theory, we would expect the image of this map to lie entirely in the $H^{i,i}(S)$ summand. In our example we will only look at the case of divisors on S, which is $Z^1(S) \to H^{1,1}(S)$. What is special about our choice of S is that we can construct six explicit algebraic cycles of codimension 1, and show by intersection theory that their images in $H^{1,1}(S)$ must be independent. This will be in opposition to our claim that $H^{1,1}(S)$ must have rank 4, and so will rule out the possibility of a Hodge decomposition that behaves just like the singular case.

So our next goal is to examine the codimension 1 algebraic cycles on S. What is special about supersingular elliptic curves is that their endomorphism algebra is rank 4 over $\mathbb{Z}$, and after tensoring with $\mathbb{Q}$ it becomes a quaternion algebra. Let 1, α, β, and γ denote abelian group generators for the endomorphism algebra (with 1 being the identity morphism).

For any morphism $f\colon C \to C$ let Γ_f denote its graph, which is a codimension one algebraic cycle on $C \times C$. We therefore have four elements of $Z^1(S)$, namely Γ_1, Γ_α, Γ_β, and Γ_γ. Add to this list the two elements $C_h = C \times *$ and $C_v = * \times C$ (and note that $C_h = \Gamma_0$, the graph of the zero homomorphism). We will calculate the matrix of intersection products of these elements, and see that it is nondegenerate. Recall that given two classes in $H^*(S)$ which come from algebraic cycles of complementary dimension, their product in $H^*(S)$ is assumed to be the intersection product. It follows that our six cycles have linearly independent images in $H^{1,1}(S)$, as was desired.

To calculate the intersection matrix we need to recall the notion of degree. For an endomorphism $f\colon C \to C$, the number of elements in the kernel is finite unless f is the zero homomorphism. One sets $\deg(f) = \#\ker(f)$ when f is nonzero, and defines $\deg(0) = 0$. One can check that $\deg(-)$ is a quadratic form on $\mathrm{End}(C)$, and it is clearly positive-definite. Note also that $\deg(fg) = \deg(f)\deg(g)$ for all $f, g \in \mathrm{End}(C)$, which shows that $\mathrm{End}(C)$ does not contain any zero divisors. So $\mathrm{End}(C)_{\mathbb{Q}}$ is a division algebra. Let $(f, g) = \frac{1}{2}[\deg(f + g) - \deg(f) - \deg(g)]$ be the symmetric bilinear form corresponding to the quadratic form $\deg(-)$.

The following identities follow from elementary geometric considerations:

$$C_v \cdot C_v = C_h \cdot C_h = 0, \quad C_v \cdot C_h = 1,$$
$$\Gamma_f \cdot C_v = 1, \quad \Gamma_f \cdot C_h = \deg(f), \quad \Gamma_f \cdot \Gamma_g = \deg(f - g).$$

In the last of these, for instance, one observes that Γ_f and Γ_g meet in points $(x, f(x))$ where $f(x) = g(x)$, these are in one-to-one correspondence with values of x for which $(f - g)(x) = 0$. The other identities involve similar considerations.

Let us define $\Gamma'_f = \Gamma_f - C_h - (\deg f)C_v$, and observe that $\Gamma'_f \cdot C_h = \Gamma'_f \cdot C_v = 0$. Then compute that

$$\Gamma'_f \cdot \Gamma'_g = \deg(f - g) - \deg(f) - \deg(g) = -2(f, g).$$

With respect to the basis Γ'_1, Γ'_α, Γ'_β, Γ'_γ, C_h, C_v the intersection matrix then has the block form

$$\begin{bmatrix} -2\mathbf{A} & \mathbf{O} \\ \mathbf{O} & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \end{bmatrix}, \tag{6.1}$$

where A is the 4×4 matrix for the form $(-,-)$ on $\operatorname{End}(C)_{\mathbb{Q}}$ with respect to the basis $1, \alpha, \beta, \gamma$. Since A is nonsingular (because the form $(-,-)$ is positive-definite), our 6×6 intersection matrix is nonsingular as well. This completes our story.

EXERCISE 6.2. It is possible to calculate the matrix A more explicitly. The form $(-,-)$ is the norm form on the quaternion algebra $\operatorname{End}(C)_{\mathbb{Q}}$, and here is where we get a bit lucky. The key is that there are not many quaternion algebras over $\mathbb{Q}$, and they have been completely classified. This exercise reviews this theory.

Quaternion algebras over a given field F are central simple algebras, and as such are matrices over a central division algebra $D \subseteq F$. Because a quaternion algebra has dimension 4, the only possibilities are that it is itself a division algebra or that it $M_2(F)$; in the latter case we say the quaternion algebra is **split** over F.

As a central simple F-algebra, a quaternion algebra A represents an element $[A]$ in the Brauer group $\operatorname{Br}(F)$. The inverse is $[A^{op}]$, but for a quaternion algebra the involution gives an isomorphism $A \cong A^{op}$. It follows that $[A]$ has order 2 in $\operatorname{Br}(F)$.

Class field theory provides a classification of division algebras over $\mathbb{Q}$ in terms of those over the completions $\mathbb{Q}_p$ (where we include the infinite prime via the convention $\mathbb{Q}_\infty = \mathbb{R}$). This is succinctly encoded via a short exact sequence

$$0 \longrightarrow \operatorname{Br}(\mathbb{Q}) \longrightarrow \bigoplus_{p \le \infty} \operatorname{Br}(\mathbb{Q}_p) \xrightarrow{\operatorname{Inv}} \mathbb{Q}/\mathbb{Z} \longrightarrow 0. \tag{6.3}$$

Here the first map is just extension of scalars along each completion $\mathbb{Q} \to \mathbb{Q}_p$, and is well-defined as a map into the direct sum because a central simple algebra will become split over all but finitely many $\mathbb{Q}_p$'s. The second map is the sum of maps $\operatorname{Inv}_p \colon \operatorname{Br}(\mathbb{Q}_p) \hookrightarrow \mathbb{Q}/\mathbb{Z}$ which associate to each division algebra over $\mathbb{Q}_p$ an invariant in $\mathbb{Q}/\mathbb{Z}$. The complete description of these maps is part of class field theory, but it's easier to describe on the level of 2-torsion because the only 2-torsion elements in $\mathbb{Q}/\mathbb{Z}$ are 0 and $\frac{1}{2}$. Given a quaternion algebra A over $\mathbb{Q}_p$, one has $\operatorname{Inv}_p(A) = 0$ if A is split over $\mathbb{Q}_p$, and $\operatorname{Inv}_p(A) = \frac{1}{2}$ otherwise. Note that since the composition of the two maps in (6.3) is zero, it follows that if A is a quaternion algebra over $\mathbb{Q}$ then the set of primes p over which A becomes non-split must be *even*.

Now consider an elliptic curve $\mathcal{E}$ defined over a field k of characteristic p. An endomorphism $f\colon \mathcal{E} \to \mathcal{E}$ will send the n-torsion points $\mathcal{E}[n]$ into itself, and the subgroups $\mathcal{E}[n] \hookrightarrow \mathcal{E}$ tend to be simple to understand. For any prime l, the Tate module $T_l(\mathcal{E})$ is defined to be the inverse limit of the tower

$$\cdots \xrightarrow{\times l} \mathcal{E}[l^3] \xrightarrow{\times l} \mathcal{E}[l^2] \xrightarrow{\times l} \mathcal{E}[l].$$

The map f induces a map $f_*\colon T_l(\mathcal{E}) \to T_l(\mathcal{E})$, and in this way one obtains a map of algebras

$$\operatorname{End}(\mathcal{E}) \to \operatorname{End}(T_l(\mathcal{E})).$$

It is not hard to argue that this is an injection.

It is known that when $l \neq p$ one has an isomorphism $E[l^e] \cong \mathbb{Z}/(l^e) \times \mathbb{Z}/(l^e)$. It follows that $T_l(\mathcal{E}) \cong \mathbb{Z}_l \times \mathbb{Z}_l$, and we obtain an injection of $\mathbb{Q}_l$-algebras

$$\operatorname{End}(\mathcal{E}) \otimes \mathbb{Q}_l \hookrightarrow \operatorname{End}_{\mathbb{Q}_l}(\mathbb{Q}_l \times \mathbb{Q}_l) = M_2(\mathbb{Q}_l).$$

From this it is apparent that the endomorphism algebra $\operatorname{End}(\mathcal{E})$ always has rank at most 4.

When the elliptic curve $\mathcal{E}$ is supersingular the rank of this endomorphism algebra is equal to 4 (this is sometimes even taken to be the definition of supersingular), and so the above map is an isomorphism: $\operatorname{End}(\mathcal{E}) \otimes \mathbb{Q}_l \cong M_2(\mathbb{Q}_l)$. In other words, $\operatorname{End}(\mathcal{E})$ is split over every prime except possibly p and ∞. It is easy to see that $\operatorname{End}(\mathcal{E})$ is *not* split over $\mathbb{R}$, as the degree form is clearly nonsingular and so $\operatorname{End}(\mathcal{E}) \otimes \mathbb{R}$ is a division algebra. It follows that $\operatorname{End}(\mathcal{E})_{\mathbb{Q}}$ is the unique quaternion algebra over $\mathbb{Q}$ which is non-split only at the primes p and ∞. To actually identify this algebra one must do a little legwork, and this is where we will leave things to the reader (but with the guide below given for structure).

Recall that a quadratic form q is said to be **isotropic** if the equation $q(x) = 0$ has nonzero solutions, and **anisotropic** otherwise. If k is a field and $a_1, \dots, a_n \in k$, write $\langle a_1, \dots, a_n\rangle$ for the quadratic form

$$q(x_1, \dots, x_n) = a_1 x_1^2 + \cdots + a_n x_n^2.$$

Over a field not of characteristic 2, every quadratic form is isomorphic to such a diagonal form. In the exercises below, let p always be an odd prime. Readers who get stuck on some parts can consult [**S1**, Chapters 2.11 and 5.6] for an excellent reference.

(a) Prove that $\mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2$ consists of four elements, represented by $1, \epsilon, p, \epsilon p$ where ϵ is any integer giving a generator for the group $\mathbb{F}_p^\times/(\mathbb{F}_p^\times)^2 \cong \mathbb{Z}/2$.

(b) Note that $\langle a\rangle = \langle ad^2\rangle$ for any $a, d \in \mathbb{Q}_p^\times$. Conclude that every 4-dimensional form over $\mathbb{Q}_p$ is isomorphic to a form $\langle s_1, s_2, s_3, s_4$ where each $s_i \in \{1, \epsilon, p, \epsilon p\}$.

(c) Prove that if $p \equiv 1 \pmod 4$, then over $\mathbb{F}_p$ the form $\langle 1, 1\rangle$ is isotropic. Use Hensel's Lemms to conclude the same for $\mathbb{Q}_p$, and deduce that $\langle x, x\rangle$ is isotropic for any $x \in \mathbb{Q}_p - \{0\}$.

(d) When $p \equiv 1 \pmod 4$, prove that every 4-dimensional form over $\mathbb{Q}_p$ is either isotropic or isomorphic to $\langle 1, \epsilon, p, \epsilon p\rangle$. Verify that the latter form is anisotropic.

(e) When $p \equiv 3 \pmod 4$, prove that $\langle 1, \epsilon\rangle$ is isotropic and that $\langle 1, 1\rangle \cong \langle \epsilon, \epsilon\rangle$. Note that then $\langle x, \epsilon x\rangle$ is isotropic for each $x \in \mathbb{Q}_p^\times$, and $\langle x, x\rangle \cong \langle \epsilon x, \epsilon x\rangle$. Using these facts, prove that every 4-dimensional quadratic form over $\mathbb{Q}_p$ is either isotropic or isomorphic to $\langle 1, 1, p, p\rangle$. Verify that the latter form is anisotropic.

(f) For $a, b \in k^{\times}$, with k a ground field, let (a, b) denote the quaternion algebra with k-basis $1, i, j, k$ defined by

$$i^2 = a, \quad j^2 = b, \quad ij = -ji, \quad ij = k.$$

This algebra has a conjugation $x \mapsto \bar{x}$ defined as usual, and the associated norm form is $N(x) = x\bar{x}$. Verify that the norm form for (a, b) is isomorphic to $\langle 1, -a, -b, ab \rangle$.

(g) Prove that when $p \equiv 1 \pmod 4$, the quaternion algebra $(-\epsilon, -p)$ is non-split at p and ∞ but splits at every other prime (the algebra splits if and only if its norm form is isotropic).

(h) Prove that when $p \equiv 3 \pmod 4$, the algebra $(1, p)$ is non-split at p and ∞ but splits at every other prime.

(i) Deduce the form of the matrix $\mathbf{A}$ in (6.1).

6.4. Newton polygons and a conjecture of Katz. Although a Hodge decomposition cannot exist for varieties in characteristic p, that doesn't mean that Hodge-like phenomena are not present. A complete survey of Hodge theory in characteristic p would take us quite a long time, and would be very technical. For now we will be content to point out some fundamental examples.

Let X be a smooth, projective variety of dimension d over a finite field $\mathbb{F}_q$, where $q = p^e$. We have talked at length about the formulas

$$\#X(\mathbb{F}_{q^m}) = 1 - [\alpha_{1,1}^m + \cdots + \alpha_{1,b_1}^m] + [\alpha_{2,1}^m + \cdots + \alpha_{2,b_2}^m] + \cdots$$

where the $\alpha_{k,1}, \ldots, \alpha_{k,b_k}$ are the eigenvalues of the Frobenius map F acting on a conjectural cohomology group $H^k(\overline{X})$. The $\alpha_{k,j}$'s are expected to be algebraic integers, and the Riemann hypothesis says that their norm should be $q^{k/2}$.

Rather than study the complex norm, we can also study the l-adic valuations of the $\alpha_{k,j}$'s for different primes l. It turns out that only $l = p$ gives something interesting, though. Indeed, by Poincaré Duality we expect that the set of eigenvalues $\{\alpha_{k,j}\}_j$ is equal to the set $\{q^d/\alpha_{2d-k,j}\}$. That is to say, for each value of j there is a j' such that $\alpha_{k,j} \cdot \alpha_{2d-k,j'} = q^d$. But if $l \neq p$ then q is a unit in $\mathbb{Z}_l$, which means that $\alpha_{k,j}$ is also a unit in $\mathbb{Z}_l$; hence its l-adic valuation is zero.

We aim to study the p-adic valuations of the $\alpha_{k,j}$'s. To this end, start by ordering the $\alpha_{k,j}$'s so that

$$\operatorname{ord}_q(\alpha_{k,1}) \leq \operatorname{ord}_q(\alpha_{k,2}) \leq \cdots \leq \operatorname{ord}_q(\alpha_{k,b_k}),$$

and also set

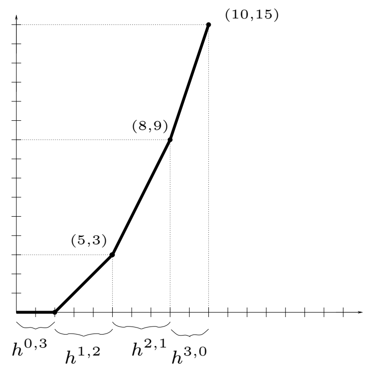
$$a_j = \operatorname{ord}_q(\alpha_{k,j}).$$

The surprising claim is that if X is the reduction of a smooth variety $\tilde{X}$ defined over a field of characteristic zero, then the numbers a_j seem to have some connection to the dimensions of the groups in the Hodge decomposition for $\tilde{X}$. The precise relationship is a bit hard to describe; the clearest approach is through the geometry of Newton polygons.

Given a smooth, projective variety Y over $\mathbb{C}$, recall that each group $H^k(Y;\mathbb{C})$ has a Hodge decomposition

$$H^k(Y;\mathbb{C}) = H^{0,k}(Y) \oplus H^{1,k-1}(Y) \oplus \cdots \oplus H^{k-1,1}(Y) \oplus H^{k,0}(Y).$$

The dimensions $h^{i,j} = \dim_{\mathbb{C}} H^{i,j}(Y)$ are called the *Hodge numbers* of Y. For a fixed k, these numbers can be represented geometrically by the following picture. Start at $(0, 0)$ and draw a line of slope 0 for $h^{0,k}$ steps along the x-axis. Picking up from the ending point, now draw a line of slope 1 for $h^{1,k-1}$ steps along the x-axis, then a line of slope 2 for $h^{2,k-2}$ steps, and so on. One gets a picture as in the following example, which shows an imagined $H^3(Y)$ where $h^{0,3} = 2 = h^{3,0}$ and $h^{1,2} = 3 = h^{2,1}$:



This picture is called the **Hodge polygon for** $H^k(Y)$. Different varieties can give rise to quite different-looking polygons, but note that the ending point of the Hodge polygon is always $(\beta_k, k \cdot \beta_k/2)$, where $\beta_k = \dim H^k(Y)$. This follows from the symmetry of the Hodge numbers $h^{i,k-i} = h^{k-i,i}$: the total vertical rise in the ith section of the Hodge polygon is $i \cdot h^{i,k-i}$, but summing this with the vertical rise in the $(k-i)$th section gives $ih^{i,k-i} + (k-i)h^{k-i,i} = kh^{i,k-i}$. So the total vertical rise of the Hodge polygon is

$$\begin{aligned}\sum_{i=0}^{k} i \cdot h^{i,k-i} &= \frac{1}{2} \cdot \left[\sum_{i=0}^{k} ih^{i,k-i} + \sum_{i=0}^{k} (k-i)h^{k-i,i} \right] \\ &= \frac{1}{2} \cdot \sum_{i=0}^{k} kh^{i,k-i} = \frac{1}{2}k \cdot \sum_{i=0}^{k} h^{i,k-i} = \frac{k\beta_k}{2}.\end{aligned}$$

Recall that we defined $a_i = \mathrm{ord}_q(\alpha_i)$. The numbers $a_1, \ldots, a_{\beta_k}$ can also be used to construct a certain polygon, this time called the **Newton polygon** for $H^k(X)$. Here one starts at $(0,0)$ and draws a line segment of slope a_1 for one step along the x-axis, then a connecting line segment of slope a_2 along one more step, then a line segment of slope a_3 for yet one more step, and so on. By the time this is done one has moved exactly β_k steps along the x-axis. In fact, if the Hard Lefschetz Theorem holds then the last point of the Newton polygon will be $(\beta_k, k \cdot \beta_k/2)$, just as for the Hodge polygon. For as we have seen previously, if $\{\alpha_1, \ldots, \alpha_k\}$ are the eigenvalues of Frobenius on $H^k(Y)$ (recorded with multiplicity), then the Hard Lefschetz Theorem implies that the two sets

$$\{\alpha_i\} \qquad \text{and} \qquad \left\{ \frac{q^k}{\alpha_i} \right\}$$

are equal. Adding up $\mathrm{ord}_q(\alpha)$ as α ranges over each of these sets, one finds that

$$\sum_{i=0}^{k} \mathrm{ord}_q(\alpha_i) = k\beta_k - \sum_{i=0}^{k} \mathrm{ord}_q(\alpha_i),$$

or $\sum_{i=0}^{k} \operatorname{ord}_q(\alpha_i) = \frac{k\beta_k}{2}$. But recall $a_i = \operatorname{ord}_q(\alpha_i)$, so this sum is also the total height of the Newton polygon.

We can now state a very interesting conjecture:

Conjecture 6.5. *Let X be a smooth, projective variety over a finite field $\mathbb{F}_q$ Then*

(a) The vertices of the Newton polygons for each cohomology group of X occur only at integral lattice points, and

(b) [Katz] If X lifts to a smooth variety $\tilde{X}$ in characteristic 0, then the Hodge polygons always lie underneath the Newton polygons. That is, if $N(x)$ and $H(x)$ are the functions whose graphs are the Newton and Hodge polygons for a cohomology group $H^k(X)$, then one has $N(x) \geq H(x)$ for all $0 \leq x \leq \beta_k(X)$.

Following Mazur [**M1**], it is nice to point out some specific consequences of the conjecture. For instance, it says that at most $h^{0,k}$ of the $\alpha_{k,j}$ eigenvalues must be p-adic units. If exactly $h^{0,k}$ of them are p-adic units, then the rest of the eigenvalues are divisible by q; and of these, at most $h^{1,k-1}$ have the property that $\alpha_{k,j}/q$ is a p-adic unit.

The conjecture of Katz can also be written algebraically, in terms of a certain inequality. It says that for any integer in the range $0 \leq t \leq \beta_k$ one has

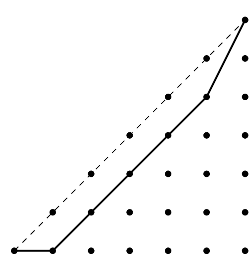
$$a_1 + \cdots + a_t \geq 0 \cdot h^{0,k} + 1 \cdot h^{1,k-1} + \cdots + j \cdot h^{j,k-j} + (j+1) \cdot (t - \beta_j) \tag{6.6}$$

where j is the unique integer such that

$$h^{0,k} + h^{1,k-1} + \cdots + h^{j,k-j} \leq t < h^{0,k} + h^{1,k-1} + \cdots + h^{j,k-j} + h^{j+1,k-j-1}.$$

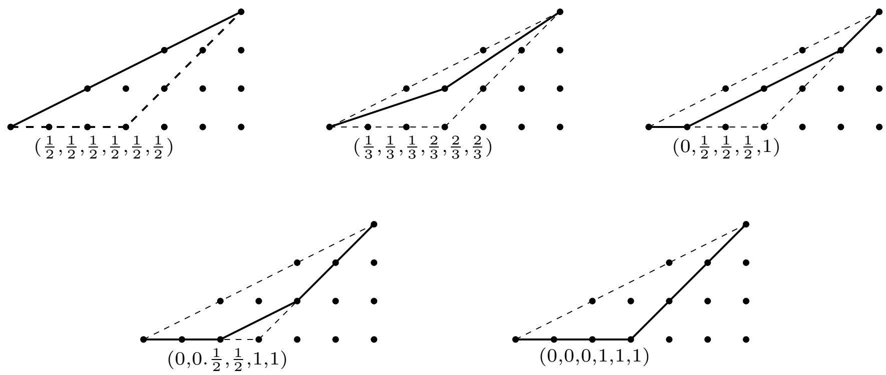
The left side of (6.6) is simply the height of the Newton polygon above the point t, whereas the right side is the height of the Hodge polygon at this same point.

Example 6.7. Let us return to our example of $X = C \times C$, where C is a supersingular elliptic curve over $\mathbb{F}_q$. Assume that q is large enough so that all the endomorphisms of C over $\overline{\mathbb{F}}_q$ are already defined over $\mathbb{F}_q$. For a lift to characteristic zero, the Hodge decomposition of $H^2(X)$ has $h^{2,0} = h^{0,2} = 1$ and $h^{1,1} = 4$, whereas we saw at the beginning of this section that the étale cohomology group $H^2(X)$ has a basis consisting of six algebraic cycles. Therefore the eigenvalues of Frobenius are q with multiplicity 6! One gets the following picture for the Newton and Hodge polygons (the Newton polygon is dashed):



Observe that this example conforms to the Katz conjecture.

Example 6.8. Let X be an algebraic curve of genus 3. Then $\dim H^1(X) = 6$, and in the case of complex varieties the Hodge decomposition would necessarily have $h^{1,0} = h^{0,1} = 3$. The Katz conjecture from above gives five possibilities for the Newton polygon of such a curve in characterisitic p, depicted in the diagrams below. One obtains these simply by considering all the possibilities for the break-points (vertices) of the polygon. Below each Newton polygon we have listed the tuple $(a_1, \ldots, a_6)$ giving the values $a_i = \mathrm{ord}_q(\alpha_i)$.



Example 6.9. Now consider the example of K3 surfaces X. For complex varieties we would have $H^2(X) = \mathbb{Z}^{22}$, and the Hodge decomposition is $h^{2,0} = h^{0,2} = 1$ and $h^{1,1} = 20$. In this case there is a very narrow range between the Hodge polygon and the line $y = x$ (which connects the origin to the endpoint $(22, 22)$ of the Hodge polygon), and the Katz conjecture says that the possible Newton polygons in characteristic p must lie in this range. The possibilities for the Newton polygon, keeping in mind Poincaré Duality and that the breakpoints occur on the integral lattice, are then

(i) There are no break points, so the polygon is simply the line segment connecting $(0, 0)$ to $(22, 22)$, or
(ii) The break points are $(0, 0)$, $(h, h-1)$, $(20-h, 19-h)$, and $(22, 22)$, for some h in the range $1 \leq h \leq 11$.

In case (ii) having $h = 11$ is actually impossible: it is known that a K3 surface must have at least one algebraic cycle of codimension one on it, which means at least one of the numbers a_i is equal to 1. So there must be some section of the Newton polygon having slope 1, and this does not occur when $h = 11$.

If we set $h = 0$ for case (i), then the number h with $0 \leq h \leq 10$ becomes a new invariant for K3 surfaces over characteristic p fields. This invariant was first investigated by Artin and Mazur.

Conjecture 6.5 has been proven, using the theory of *crystalline cohomology.* The story can be explained as follows. For varieties over a field of characteristic p, étale cohomology with $\mathbb{Q}_p$ coefficients turns out not to be very well behaved. Perhaps taking some hints from ealier work of Dwork, Manin, and Monsky-Washnitzer, Grothendieck envisioned a new theory with $\mathbb{Q}_p$ coefficients constructed by adapting de Rham theory into characteristic p. This theory was developed by Berthelot in his thesis. The first part of Conjecture 6.5, the fact that the vertices of the Newton polygon occur at integral lattice points, becomes a triviality: it comes about as a general property of the types of objects that arise in the crystalline theory. The second part of Conjecture 6.5, the Katz conjecture, is more subtle. It was first proved by Mazur [**M1, M2**] under some mild assumptions on the varieties X, and later in complete generality by Ogus [**BO**].

We will not say more about crystalline cohomology at the moment. The interested reader can look ahead to Chapter ????.

7. The Tate conjecture

Recall, once again, the formulas

$$\#X(\mathbb{F}_{q^m}) = 1 - [\alpha_{1,1}^m + \cdots + \alpha_{1,b_1}^m] + [\alpha_{2,1}^m + \cdots + \alpha_{2,b_2}^m] + \cdots \tag{7.1}$$

We saw in Chapter 1 that if X is projective space or a Grassmannian then the $\alpha_{i,j}$'s in this formula are very simple: they are all powers of q. Cohomologically, the reason is that if $u \in H^{2k}(X)$ is an algebraic cycle then u is an eigenvector of Frobenius with eigenvalue q^k. For projective spaces and Grassmannians every cohomology class is algebraic, so every eigenvalue of Frobenius is an integral power of q.

The Tate conjecture is a kind of converse to the above: loosely phrased, it says that *all* of the integral powers of q among the $\alpha_{i,j}$'s come from algebraic cycles.

To state the conjecture more carefully, recall that $A^i(X) \subseteq H^{2i}(X)$ denotes the $\mathbb{Q}$-vector space spanned by the fundamental classes of algebraic cycles. This is not even known to be finite-dimensional, although conjecturally it should be. The Tate conjecture can be stated in either of the following equivalent ways:

(1) The dimension of $A^k(X)$ is equal to the number of q^k's appearing among the $\alpha_{i,j}$'s in the formula (7.1);
(2) The dimension of $A^k(X)$ is equal to the order of the pole of $\zeta_X(s)$ at the point $s = q^i$.

EXAMPLE 7.2. Let C be a supersingular elliptic curve over $\mathbb{F}_p$. Then the eigenvalues of F on $H^1(C)$ must be $i\sqrt{p}$ and $-i\sqrt{p}$. On $H^0(C)$ and $H^2(C)$ the eigenvalues are of course 1 and p, respectively. By the Künneth Theorem the eigenvalues of F on $H^i(C \times C)$ are

H^0	H^1	H^2	H^3	H^4
1	$i\sqrt{p}\,(2),\quad -i\sqrt{p}\,(2)$	$p\,(4),\quad -p\,(2)$	$ip^{3/2}\,(2),\quad -ip^{3/2}\,(2)$	p^2

where the numbers in the parentheses represent the multiplicity of the eigenvalue. According to the Tate conjecture, the subspace of $H^2(\overline{C \times C})$ spanned by the algebraic cycles of $C \times C$ should be 4-dimensional.

This might be slightly confusing, since we have previously seen that all the elements of $H^2(\overline{C \times C})$ should be algebraic! While this is true, it only means that each class can be represented as an algebraic cycle over some extension of $C \times C$ from the base field $\mathbb{F}_p$ to some larger field. That is to say, not all the algebraic cycles giving elements of $H^2(\overline{C \times C})$ will be defined over $\mathbb{F}_p$. Indeed, only four of them will be.

If we base extend C to the field $\mathbb{F}_{p^2}$ and consider the corresponding Frobenius map (which will be the square of the Frobenius considered above), then the eigenvalues on $H^2(\overline{C \times C})$ are now

H^0	H^1	H^2	H^3	H^4
1	$-p(4)$	$p^2\,(6)$	$-p^3(4)$	p^4

The Tate conjecture now predicts that all the cohomology classes of $H^2(\overline{C \times C})$ are spanned by algebraic cycles defined over $\mathbb{F}_{p^2}$. ????

8. The Weil conjectures for abelian varieties

THEOREM 8.1. *Let A be an abelian variety of dimension d. Then* $\dim H^1(A) \leq 2d$*, and if* $\dim H^1(A) = 2d$ *then there is an isomorphism of rings* $H^*(A) \cong \bigwedge^* \big[H^1(A)\big]$.

PROOF. Write

$$\nabla\colon H^*(A) \to H^*(A^{2d}) \cong \bigotimes_{i=1}^{2d} H^*(A)$$

for the map on cohomology induced by the (iterated) multiplication map $A^{2d} \to A$. Note that this is a ring map. For $x \in H^*(A)$, let $x(i)$ denote $1\otimes\cdots\otimes 1\otimes x\otimes 1\otimes\cdots\otimes 1$, with the x appearing in the ith factor. It follows easily that for any $x \in H^*(A)$ one has

$$\nabla(x) = [x(1) + x(2) + \cdots + x(2d)] + \sum y_{i_1} \otimes \cdots \otimes y_{i_{2d}}$$

where the terms inside the sum have each y_i homogeneous and at least two of the y_i's of positive degree. In particular, note that if $x \in H^1(A)$ then $\nabla(x) = x(1) + \cdots + x(2d)$.

Let $x_1, \ldots, x_m \in H^1(A)$. Then $\nabla(x_1 \cdots x_m) = \nabla(x_1) \cdots \nabla(x_m)$. In multidegree $(1, 1, \ldots, 1)$ the right hand side is

$$\sum_\sigma x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(2d)}, \tag{8.2}$$

where σ ranges over all the permutations of $2d$ letters. If $x_1, \ldots, x_m$ are linearly independent then the terms in the sum (8.2) are also linearly independent, and so this sum is nonzero. Hence $\nabla(x_1 \cdots x_m) \neq 0$, and therefore $x_1 \cdots x_m \neq 0$ as well. But since $H^i(A) = 0$ for $i > 2d$, it must be that $m \leq 2d$. So we have shown that $\dim H^1(A) \leq 2d$.

Now suppose that $\dim H^1(A) = 2d$, and let $x_1, \ldots, x_{2d}$ be a basis. Since $H^*(A)$ is graded commutative, we have the evident algebra map

$$f\colon \bigwedge^*[H^1(A)] \to H^*(A).$$

By the previous paragraph we know that $x_1 x_2 \cdots x_{2d}$ is nonzero, hence f is an isomorphism in degree $2d$.

Suppose that α is a nonzero homogeneous element of the domain of f. There exists a β in $\bigwedge^*[H^1(A)]$ such that $\alpha \wedge \beta = x_1 \wedge \cdots \wedge x_{2d}$, therefore $f(\alpha)f(\beta) = f(\alpha \wedge \beta) \neq 0$. In particular, this implies $f(\alpha) \neq 0$, and so f is injective.

By construction f is surjective in dimension 1. Let $z \in H^k(A)$ and assume by induction that f has been proven to be surjective in all dimensions smaller than k. Start with the very silly equation $x_1 \cdots x_{2d} z = 0$ and apply μ^*, where $\mu\colon A \times A \to A$ is the multiplication:

$$\begin{aligned} 0 &= \mu^*(x_1)\mu^*(x_2) \cdots \mu^*(x_{2d})\mu^*(z) \\ &= [x_1 \otimes 1 + 1 \otimes x_1] \cdots [x_{2d} \otimes 1 + 1 \otimes x_{2d}] \cdot \Big[z \otimes 1 + 1 \otimes z + \sum_i y_i' \otimes y_i''\Big] \end{aligned}$$

where the degrees of y_i' and y_i'' are positive and strictly less than k.

Multiply out the above product and group together all terms having bidegree $(2d, k)$. Such terms can appear in the above product in three ways: as

$$\begin{array}{ll} (1) & (x_1 x_2 \cdots x_{2d}) \otimes z, \\ (2) & (x_{j_1} x_{j_2} \cdots x_{j_{2d-k}} z) \otimes (x_{m_1} x_{m_2} \cdots x_{m_k}), \quad \text{or as} \\ (3) & (x_{j_1} x_{j_2} \cdots x_{j_r} y_i') \otimes (x_{m_1} x_{m_2} \cdots x_{m_s} y_i''), \end{array}$$

where in each of the last two lines the j's and the m's are disjoint sets of indices whose union is $\{1, \ldots, 2d\}$. Our equation tells us that the sum of all these terms is zero.

But note that $H^{2d}(A)$ is one-dimensional, so all elements of this vector space are multiples of $x_1 x_2 \cdots x_{2d}$ (which we have already proven is nonzero). This applies to the terms on the left of the above tensor symbols. The terms $x_{m_1} \cdots x_{m_k}$ and $x_{m_1} \cdots x_{m_s} y_i''$ are all in the image of f, in the latter case by our induction hypothesis because $|y_i''| < k$. So each of the tensors of types (2) and (3) has the form $u_q(x_1 \ldots x_{2d}) \otimes f(v_q)$ for some $u_q \in E$ and some v_q in the domain of f. We therefore obtain

$$0 = \big[(x_1 \ldots x_{2d}) \otimes z\big] + \sum_q u_q(x_1 \ldots x_{2d}) \otimes f(v_q) = (x_1 \ldots x_{2d}) \otimes \Big[z + \sum_q f(u_q v_q)\Big].$$

It follows at once that $z = -\sum_q f(u_q v_q)$, and hence z is in the image of f. $\square$

Summary.

???

Part 2

Machinery

CHAPTER 4

Introduction to étale cohomology

This chapter is still in progress!

In this chapter our goal is to give an intuitive look at étale cohomology and étale homotopy theory. This is a subject which is hard to explain all at once, as there is more than one important idea lurking behind the scenes. Our discussion will be divided into roughly the following areas:

- The basic idea: homotopy approximations
- Étale maps and coverings
- Systems of spaces
- Hypercovers
- Rigid hypercovers and the étale homotopy type
- Sheaf cohomology.

Now, we should remark that this is not quite the historical approach taken by Grothendieck. Grothendieck's technical skill allowed him to *begin* with sheaf cohomology, and to sweep most of the other topics under the rug. In some sense this was a necessity at the time, because the machinery of abstact homotopy theory was not developed enough to handle some of these other topics. But even though one *can* start the story with sheaf cohomology, and in that way package lots of the technicalities into one bundle, I think this results in a loss of intuition about what's really happening. So in our development we are going to be the turtle rather than the hare. Our discussion in this chapter owes quite a bit to [**Su1**], which we highly recommend.

1. Overview of some key points

Before jumping into our main discussion we will give a brief overview. This overview, however, will probably not make much sense until one has read the rest of the chapter! Still, it seems best to begin with a broad outline, where we call attention to certain key ideas which will be important. Everything we say here will be discussed in more detail in the coming sections.

1.1. The étale topological type. The main idea in étale homotopy theory is that to every scheme X one attaches a small category I (which depends on X) and a diagram

$$\mathrm{Et}_X \colon I \to s\mathcal{S}et.$$

The simplicial sets in this diagram are thought of as 'approximations' to the homotopy type of X, and the diagram itself will often be referred to as a 'system of approximations'. In the case where X is a scheme over $\mathbb{C}$ these really *are* approximations to the classical homotopy type of $X(\mathbb{C})$, in a way that can be made precise. In the case where X is defined over another field, perhaps of characteristic

p, there is no "classical homotopy type" for us to compare things to—instead all we have are these 'approximations', and étale homotopy theory is really about learning what one can do with them.

The diagram Et_X is sometimes called the "étale topological type" of X, or the "étale realization of X". There are a few things we should say about it up front. First of all, the category I will be cofiltered. This means that for any two objects i and j in I, there is a third object k together with maps $k \to i$ and $k \to j$; also, if $i \rightrightarrows j$ are two maps in I then there is an object k and a map $k \to i$ such that the two composites $k \rightrightarrows j$ are equal. Such indexing categories are good for taking inverse limits, for reasons we will not describe right now. But one should think of diagrams indexed by I as special kinds of inverse limit systems.

Each space in the diagram Et_X, by itself, is not a very good approximation to X—it does not have much useful information about X in it. But taken altogether, as a system, there *is* some very useful information about X; it is essentially encoded in the "limit" of the system, although one has to be very careful how one interprets that. If one were to actually take the limit, or even the homotopy limit, it turns out that lots of important information is thrown away. One of the arts of this subject is learning how to extract that important information.

This idea of having a system of approximations, where the useful information is somehow "in the limit", is probably a bit strange. In Section 2 we will discuss a familiar topological context where such things occur naturally, and hopefully that will make the situation clearer.

1.2. Homotopy invariants of systems. Once we have defined Et_X, our goal will be to extract useful information from it. We will need to talk about cohomology theories for systems, for example. Let us introduce a bit more language. A *pro-space* is a diagram $I \to s\mathcal{S}et$ in which the indexing category I is cofiltered. One can make a sensible category out of such objects (where the indexing categories are allowed to vary), and this category is denoted pro-s$\mathcal{S}$et.

The singular cohomology of a pro-space $W\colon I \to s\mathcal{S}et$, with coefficients in an abelian group A, is defined simply as

$$H^n(W; A) = \operatorname*{colim}_{i \in I^{op}} H^n(W_i; A).$$

This seems simple enough, but already note that it is different from $H^n(\lim_I W; A)$ or $H^n(\operatorname{holim}_I W; A)$—an indication that taking a limit or homotopy limit of W would have been the wrong thing to do.

Adapting other cohomology theories to give invariants of systems turns out to be more complicated. If E is a cohomology theory where only finitely many of the coefficient groups $E^*(pt)$ are nonzero, then one can use the same definition:

$$E^n(W) = \operatorname*{colim}_{i \in I^{op}} E^n(W).$$

But for K-theory, for example, we have to define

$$K^{-n}(W) = \pi_n \left[\operatorname*{holim}_{k \in \mathbb{N}} \operatorname*{hocolim}_{i \in I^{op}} \operatorname{Map}(W_i, P_k(\mathbb{Z} \times BU)) \right] \tag{1.3}$$

for $n > 0$. Here $P_k(-)$ is the nth Posntikov sections functor. The complexity of (1.3) is daunting, particularly the presence of *two* limits (which cannot be commuted). We will see later why this formula is "well-behaved", whereas simpler formulas like $\operatorname{colim}_i K^n(W_i)$ are not.

The *étale cohomology* of a scheme X will be defined to be the singular cohomology of the pro-space Et_X. Likewise, the *étale K-theory* of X will be defined to be the K-theory of Et_X. We are skipping some complications, in that one needs to be able to treat cohomology not only with constant coefficients but with twisted coefficients as well—and for that one needs a little extra work. But the basic picture we've presented is valid.

1.4. Homotopy invariants and model categories. In some sense the right way to look at the above invariants is via model category theory. It is possible to set up a model category structure on pro-s$\mathcal{S}$et where the above invariants arise as homotopy classes of maps. Note that any space Z can be regarded as a pro-space by having the indexing category I be the trivial category with one object and an identity map. Write cZ for Z regarded as a pro-space. We will ultimately see that

$$[W, cK(A,n)] \cong H^n(W;A)$$

and that

$$[\Sigma^n W, c(\mathbb{Z}\times BU)] \cong K^{-n}(W).$$

So we find that studying homotopical invariants of pro-spaces is really the same as studying $\mathrm{Ho}\,(\text{pro-s}\mathcal{S}\text{et})$. The image of Et_X under the canonical map

$$\text{pro-s}\mathcal{S}\text{et} \to \mathrm{Ho}\,(\text{pro-s}\mathcal{S}\text{et})$$

is called the *étale homotopy type* of X.

We note that there is a map $\mathrm{Ho}\,(\text{pro-s}\mathcal{S}\text{et}) \to \text{pro}-\mathrm{Ho}\,(s\mathcal{S}et)$. The image of Et_X in $\text{pro}-\mathrm{Ho}\,(s\mathcal{S}et)$ is sometimes called the *classical étale homotopy type* of X. This was what was originally defined by Artin and Mazur [**AM**], at the very beginnings of the subject. But in modern timees it is acknowledged that it is better to work in $\mathrm{Ho}\,(\text{pro-s}\mathcal{S}\text{et})$, or even in the model category pro-s$\mathcal{S}$et itself—the theory is tighter and more robust when developed in those settings.

2. Topological perspectives

Let X be a topological space and let $\mathcal{U} = \{U_\alpha\}$ be an open cover of X indexed by a set A. The **Čech complex** of this open cover is the simplicial space $\check{C}(\mathcal{U})$ depicted below:

$$\coprod U_{\alpha_0} \leftleftarrows \coprod U_{\alpha_0\alpha_1} \;\substack{\leftarrow\\[-2pt]\leftarrow\\[-2pt]\leftarrow}\; \coprod U_{\alpha_0\alpha_1\alpha_2}\cdots$$

Here $U_{\alpha_0\cdots\alpha_n} = U_{\alpha_0}\cap\cdots\cap U_{\alpha_n}$, and the face maps are obtained by omitting indices—we have chosen not to draw the degeneracies for typographical reasons. The coproduct in level n is indexed by all $(n+1)$-tuples $(a_0,\ldots,a_n)\in A^{n+1}$.

Note that $\check{C}(\mathcal{U})$ is an augmented simplicial space, via the map $\coprod_\alpha U_\alpha \to X$. This augmentation can also be regarded as a map of simplicial spaces $\check{C}(\mathcal{U}) \to cX$, where cX is the constant simplicial space with X in every dimension. We will usually write just "X" rather than "cX", by abuse.

Given a simplicial space Z_*, we can form its geometric realization $|Z|$. We can also regard Z as a functor $\Delta^{op} \to \mathcal{T}op$ and construct its homotopy colimit $\mathrm{hocolim}\, Z$. Under reasonable hypotheses on Z_* these two objects will be weakly equivalent. In this chapter we will usually phrase results in terms of the homotopy colimit, but readers should note that the geometric realization gives a smaller model for what is basically the same thing.

Our map of simplicial spaces $\check{C}(\mathcal{U}) \to cX$ gives rise to a map $\operatorname{hocolim} \check{C}(\mathcal{U}) \to \operatorname{hocolim}(cX)$, and there is a canonical map $\operatorname{hocolim}(cX) \to \operatorname{colim}(cX) \cong X$. We therefore have a natural map of spaces

$$\operatorname{hocolim} \check{C}(\mathcal{U}) \to X.$$

THEOREM 2.1. *Let $\{U_\alpha\}$ be an open cover of a topological space X. Then $\operatorname{hocolim} \check{C}(\mathcal{U}) \to X$ is a weak equivalence. If the U_α's and their iterated intersections are all cofibrant, then $|\check{C}(\mathcal{U})| \to X$ is a weak equivalence.*

The above result was essentially proven by Segal [**S1**] in the case where there exists a partition of unity subordinate to the cover $\mathcal{U}$. For a modern proof that doesn't require this condition, see [**DI**].

REMARK 2.2. If a well-ordering of the set A is chosen, then one can form the **ordered Čech complex $\check{C}^o(\mathcal{U})$**. This is the simplicial space

$$[n] \mapsto \coprod_{\alpha_0 \le \cdots \le \alpha_n} U_{\alpha_0 \cdots \alpha_n}$$

There is a map of simplicial spaces $\check{C}^o(\mathcal{U}) \to \check{C}(\mathcal{U})$, and this always induces a weak equivalence of homotopy colimits. See [**DI**, Proposition 2.6].

The advantage of the ordered Čech complex is that it is quite a bit smaller; for instance, when the open cover is finite then the ordered Čech complex has only finitely many non-degenerate pieces. The regular Čech complex tends to have infinitely many such pieces, even for a simple two-fold cover $\{U_0, U_1\}$. In this case, the iterated intersections $U_0 \cap U_1 \cap U_0 \cap \cdots \cap U_0 \cap U_1$ are all nondegenerate.

We now consider several examples of Theorem 2.1. The examples deal with the ordered Čech complex, as it is easier to handle.

EXAMPLE 2.3. Let $X = S^1$, thought of as the unit complex numbers. Let

$$U = \left\{ e^{i\theta} \,\middle|\, -\frac{1}{10} < \theta < \frac{2\pi}{3} + \frac{1}{10} \right\}, \quad V = \left\{ e^{i\theta} \,\middle|\, \frac{2\pi}{3} - \frac{1}{10} < \theta < \frac{4\pi}{3} + \frac{1}{10} \right\},$$
$$W = \left\{ e^{i\theta} \,\middle|\, \frac{4\pi}{3} - \frac{1}{10} < \theta < 2\pi + \frac{1}{10} \right\}.$$

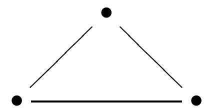
Then $\{U, V, W\}$ is an open cover of X. The nondegenerate terms in the Čech complex can be drawn as

$$\begin{tikzcd} U\cap V \arrow[d, "d_1"'] \arrow[dr, "d_0"] & U\cap W \arrow[dl, "d_1"] \arrow[dr, "d_0"] & V\cap W \arrow[dl, "d_1"'] \arrow[d, "d_0"] \\ U & V & W. \end{tikzcd}$$

Using the fact that all of U, V, W, $U \cap V$, $U \cap W$, and $V \cap W$ are contractible, the realization of $\check{C}(\mathcal{U})$ is weakly equivalent to the realization of

$$\begin{tikzcd} * \arrow[d, "d_1"'] \arrow[dr, "d_0"] & * \arrow[dl, "d_1"] \arrow[dr, "d_0"] & * \arrow[dl, "d_1"'] \arrow[d, "d_0"] \\ * & * & *. \end{tikzcd}$$

But this realization is just the space



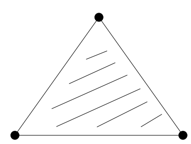
which of course is homotopy equivalent to our original X.

EXAMPLE 2.4. Now assume X is the disk $\{z \in \mathbb{C} \mid 1 \geq |z|\}$. Let $\{U, V, W\}$ be the open cover obtained by dividing the disk into three sectors of $\frac{2\pi}{3}$ radians and then 'fattenting up' the sectors a tiny bit to give open sets. (So upon intersecting with S^1, this becomes an open cover like the one considered in the last example).

For this case the nondegenerate terms in the (ordered) Čech complex look like

$$\begin{tikzcd}
 & U \cap V \cap W \arrow[dl, "d_0"'] \arrow[d, "d_1"] \arrow[dr, "d_2"] & \\
U \cap V \arrow[d, "d_1"'] \arrow[dr, "d_0"] & U \cap W \arrow[dl, "d_1"] \arrow[dr, "d_0"] & V \cap W \arrow[dl, "d_1"'] \arrow[d, "d_0"] \\
U & V & W.
\end{tikzcd}$$

Once again, all the spaces appearing in the diagram are contractible, so up to weak equivalence the realization is the same as the realization of the corresponding diagram where all the spaces have been replaced with points. The realization is therefore the space



which again is homotopy equivalent to our original X.

EXAMPLE 2.5. Generalizing the previous example, suppose a space X has an open cover $\mathcal{U}$ consisting of n open sets such that each iterated intersection is contractible. Then the simplicial space $\check{C}(\mathcal{U})$ is weakly equivalent to the simplicial set Δ^n (regarded as a simplicial space which is discrete in every dimension), and therefore the geometric realization is contractible.

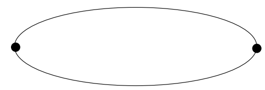
EXAMPLE 2.6. What happens when the iterated intersections $U_{\alpha_1 \ldots \alpha_k}$ are not necessarily contractible? Consider again $X = S^1$, this time with the open cover $\{U, V\}$ where

$$U = \left\{ e^{i\theta} \,\middle|\, -\frac{1}{10} < \theta < \pi + \frac{1}{10} \right\}, \qquad V = \left\{ e^{i\theta} \,\middle|\, \pi - \frac{1}{10} < \theta < 2\pi + \frac{1}{10} \right\}.$$

Then U and V are contractible, but $U \cap V$ is homotopy equivalent to S^0. The nondegenerate pieces of the ordered Čech complex now look like

$$\begin{tikzcd}
 & S^0 \arrow[dl] \arrow[dr] & \\
* & & *
\end{tikzcd}$$

and the geometric realization gives the space



.

Note again that this is homotopy equivalent to our original X.

2.7. Čech approximations. Given a space X with open cover $\mathcal{U}$, the Čech complex represents a kind of 'fattening up' of X. Its realization is a space which is weakly equivalent to X, but which is generally much bigger. However, we have seen in the above examples that if all the iterated intersections $U_{\alpha_0\cdots\alpha_n}$ are empty or contractible then we can replace the Čech complex by a smaller model—a simplicial *set*—whose realization still has the correct homotopy type. One can think of this as distilling the information in the Čech complex.

We will describe two slightly different approaches to this distilling process. Let $D\check{C}(\mathcal{U})$ denote the simplicial set obtained from $\check{C}(\mathcal{U})$ by replacing each nonempty intersection $U_{\alpha_0\cdots\alpha_n}$ with a single point. So

$$D\check{C}(\mathcal{U})_n = \{(\alpha_0, \alpha_1, \ldots, \alpha_n) \,|\, U_{\alpha_0\cdots\alpha_n} \neq \emptyset\}$$

and the face and degeneracy maps come from deleting or repeating indices. Note that there is a map $\check{C}(\mathcal{U}) \to D\check{C}(\mathcal{U})$, and if all the iterated intersections of $\mathcal{U}$ are empty or contractible then this is a levelwise weak equivalence.

The simplicial set $D\check{C}(\mathcal{U})$ is sometimes called the **Čech nerve** of the cover $\mathcal{U}$.

Remark 2.8. There is another construction which one might be tempted to call the "Čech nerve". Consider the subcategory $\mathrm{cat}(\mathcal{U})$ of $\mathcal{T}op$ consisting of all the iterated intersections $U_{\alpha_0\cdots\alpha_n}$ and the inclusion maps between them. The nerve of this category is related to $D\check{C}(\mathcal{U})$, although they are not identical. The reason is that $D\check{C}(\mathcal{U})$ really depends on the indexing set A, whereas the nerve of $\mathrm{cat}(\mathcal{U})$ does not. Some information about how these two constructions are related may be obtained from [**DI**, Cor. 3.3].

Now we describe a second way of obtaining a simplicial set from the Čech complex. For any space W, let $\pi_0(W)$ be the usual set of path components but given the quotient topology with respect to the map $W \to \pi_0(W)$. Note that for any 'reasonable' space this topology will be discrete.

Let $\pi_0\check{C}(\mathcal{U})$ denote the simplicial space obtained by applying $\pi_0(-)$ to every level of $\check{C}(\mathcal{U})$. That is, $\pi_0\check{C}(\mathcal{U})$ is the simplicial space

$$[n] \mapsto \coprod_{\alpha_0,\ldots,\alpha_n} \pi_0(U_{\alpha_0} \cap \cdots \cap U_{\alpha_n}).$$

Then once again we have a map $\check{C}(\mathcal{U}) \to \pi_0\check{C}(\mathcal{U})$. Note that if each of the iterated intersections is both 'reasonable' and homotopy discrete (i.e., weakly equivalent to a discrete space) then $\pi_0\check{C}(\mathcal{U})$ is actually a simplicial *set* and our map $\check{C}(\mathcal{U}) \to \pi_0\check{C}(\mathcal{U})$ is an objectwise weak equivalence.

Finally, we need a brief remark on what happens when one has two open covers and wants to compare their Čech complexes. Recall that our open covers are all really *indexed* open covers—that is, there is an indexing set A and for each $\alpha \in A$ we are given an open set $U_\alpha \subseteq X$. If $\{V_\beta\}_{\beta\in B}$ is another open cover of X, then by a **map of open covers** $\mathcal{U} \to \mathcal{V}$ we mean a function $f\colon A \to B$ together with maps $U_\alpha \to V_{f(\alpha)}$ for every $\alpha \in A$. In this situation one also says that $\mathcal{U}$ is a **refinement** of $\mathcal{V}$.

Given a map of covers $\mathcal{U} \to \mathcal{V}$, there is a naturally associated map of simplicial spaces $\check{C}(\mathcal{U}) \to \check{C}(\mathcal{V})$. One then obtains an induced map of simplicial sets $\pi_0\check{C}(\mathcal{U}) \to \pi_0\check{C}(\mathcal{V})$.

If $\mathcal{U}$ and $\mathcal{V}$ are any two open covers of X, note that we may form a new open cover $\mathcal{U} \cap \mathcal{V}$ by considering the set $\{U_\alpha \cap V_\beta\}_{(\alpha\in A,\beta\in B)}$. There are evident maps $\mathcal{U} \cap \mathcal{V} \to \mathcal{U}$ and $\mathcal{U} \cap \mathcal{V} \to \mathcal{V}$ giving rise to comparison maps of Čech complexes

$$\check{C}(\mathcal{U}) \leftarrow \check{C}(\mathcal{U} \cap \mathcal{V}) \to \check{C}(\mathcal{V}).$$

2.9. Systems of Čech approximations. At the expense of a few white lies, we can now give the main idea behind étale homotopy types.

If X is a topological space, let I_X be the category of open covers of X, where the maps are refinements. Consider the functor

$$\pi_0\check{C} \colon I_X \to s\mathcal{S}et$$

sending an open cover $\mathcal{U}$ to $\pi_0\check{C}(\mathcal{U})$. We will think of each simplicial set $\pi_0\check{C}(\mathcal{U})$ as a 'combinatorial approximation' to the homotopy type of X, and the functor $\pi_0\check{C}$ should be thought of as a *system* of combinatorial approximations. For an arbitrary open cover $\mathcal{U}$, the homotopy type of $\pi_0\check{C}(\mathcal{U})$ is probably not very close to the homotopy type of X—but our intuition suggests that by refining $\mathcal{U}$ we may obtain better and better approximations. In particular, it follows from Theorem 2.1 that if we can refine $\mathcal{U}$ to an open cover $\mathcal{U}'$ in which the k-fold intersections are all homotopy discrete for $0 \le k \le N$, then $\pi_0\check{C}(\mathcal{U}')$ has the same N-type as X. If we can produce such a refinement for every N (which is not at all clear), this suggests that the system of combinatorial approximations in some sense "converges" to the homotopy type of X.

The goal of étale homotopy theory is to repeat this kind of construction but starting with an algebraic variety rather than a topological space. To each algebraic variety X—not necessarily defined over the complex numbers, or even over a characteristic zero field—we will associated a category I_X and a functor

$$\mathrm{Et}_X \colon I_X \to s\mathcal{S}et.$$

This will play the role of a 'system of combinatorial approximations', from which we can extract homotopy invariants. This extraction will involve some kind of "limiting" process. The rest of this chapter will essentially be spent showing how to define I_X, how to define this functor, and precisely what this limiting process is.

It has perhaps already occured to the reader that there is in fact an obvious way to define such a system, using Zariski open covers; so perhaps we should explain right away why this obvious method doesn't work.

Let X be a scheme. For each Zariski open cover $\{U_\alpha\}$ of X, indexed by a set A, we can again form the associated Čech complex $\check{C}(\mathcal{U})$; this is now a simplicial scheme. If $\pi_0(W)$ denotes the set of connected components of W, for any scheme W, then we can also form the simplicial set $\pi_0\check{C}(\mathcal{U})$. So an obvious thing to do is to take I_X to be the category of Zariski open covers of X, and to consider the functor $\pi_0\check{C} \colon I_X \to s\mathcal{S}et$ just as we did for topological spaces.

For this to be a sensible thing to look at, we have to hope that as one takes covers by smaller and smaller Zariski open sets, the simplicial set $\pi_0\check{C}(\mathcal{U})$ becomes a better and better approximation to an interesting homotopy type. Unfortunately, this is not the case. The problem is that Zariski open sets are all very big (compared to the open sets one has in the classical topology for a variety over $\mathbb{C}$). In particular, if X is irreducible then any finite collection of Zariski open sets have a nontrivial intersection. This means that when forming the space $\pi_0\check{C}(\mathcal{U})$ one is in a situation like Example 2.4 or Example 2.5: the fact that each possible iterated intersection is nontrivial forces $|\pi_0\check{C}(\mathcal{U})|$ to be contractible. In particular, this is true no matter how much we refine the cover $\mathcal{U}$.

The conclusion is that this particular method for attaching homotopy invariants to a scheme X is hopeless.

Luckily, this method can in some ways be saved. The most important idea is to replace our open covers by a more general kind of 'cover', and we will start to explore this in the next section. On top of this, there are some delicate technical problems involved in choosing a nice enough category I_X, and even more problems associated with whether Čech complexes can be used to give 'good enough' approximations. All of this will be explained in more detail later.

3. Rigid open covers and generalized Čech complexes

In the last section we saw how to attach to any topological space X a system of Čech approximations to X, indexed by the category of open covers of X. In this section our goal is to modify this construction in two ways. First, we will tweak the indexing category just a little in order to make it cofiltered (see below for the definition). The reasons for making this modification are technical, and in the end probably unsatisfying—but things seem to work best if we can get ourselves into the cofiltered setting. The notion of a "rigid" open cover is what does this for us.

The second goal is to replace open covers with a much more general notion. This point is much more important, and is in some sense the one main insight that makes the whole étale machinery work.

3.1. Filtered and cofiltered categories. We begin with the basic definitions.

Definition 3.2. *A **filtered category** is a small category I satisfying the following two conditions:*

(a) For all objects i,j in I, there exists an object k in I and maps $i \to k$ and $j \to k$.
(b) For all objects i, j in I and all maps $f, g\colon i \to j$, there is an object k in I and a map $u\colon j \to k$ such that $uf = ug$.

The following proposition brings together most of the things we will routinely use about diagrams indexed by filtered categories.

Proposition 3.3. *Let I be a filtered category.*

(a) Suppose $A\colon I \to \mathcal{A}b$ is a diagram of abelian groups. Then an element $x \in A_i$ maps to zero under $A_i \to \operatorname{colim}_I A$ if and only if there exists an object j in I and a map $i \to j$ such that x maps to zero under $A_i \to A_j$.
(b) Suppose $A\colon I \to \mathrm{Ch}(\mathbb{Z})$ is a diagram of chain complexes. Then for all $n \in \mathbb{Z}$ the natural map $\operatorname{colim}_i H_n(A_i) \to H_n(\operatorname{colim}_I A)$ is an isomorphism.
(c) Suppose $A\colon I \to s\mathcal{S}et_*$ is a diagram of pointed simplicial sets. Then for all $n \geq 0$, the natural map $\operatorname{colim}_i \pi_n(A_i, *) \to \pi_n(\operatorname{colim} A_i, *)$ is an isomorphism. In addition, the map $\operatorname{hocolim} A_i \to \operatorname{colim} A_i$ is a weak equivalence.
(d) Suppose $A\colon I \to \mathcal{T}op$ is a diagram of pointed spaces. Then the natural map $\operatorname{colim}_i \pi_k(A_i, *) \to \pi_k(\operatorname{hocolim}_i A_i, *)$ is an isomorphism, for all $k \geq 0$.

Sketch of proof. Part (a) is routine, and part (b) is an easy consequence of (a).

To prove (c) one needs two observations. First, the result is easy when all the A_i's are fibrant (in which case $\operatorname{colim}_I A$ is also fibrant). Second, the functor Ex^1 preserves filtered colimits, and hence Ex^∞ does as well.

Finally, (d) is deduced from (c) using the Quillen equivalence between $\mathcal{T}op$ and $s\mathcal{S}et$. $\square$

Example 3.4. Both the conditions for being filtered are necessary for the above properties to hold. For example, let $\mathcal{C}$ be the co-equalizer category $\mathbf{0} \rightrightarrows \mathbf{1}$ consisting of two objects and two non-identity maps, as shown. Then $\mathcal{C}$ satisfies the first condition for being filtered, but not the second. Consider the diagram of abelian groups $\mathbb{Z} \rightrightarrows \mathbb{Z}$ where the top map is multiplication by 2 and the bottom is multiplication by 3. Then the colimit of this diagram is zero, but it is not true that every element in the diagram maps to zero somewhere else in the diagram. Thus, part (a) of the proposition does not hold for diagrams indexed by $\mathcal{C}$.

A category I is said to be **co-filtered** if I^{op} is filtered. If $\mathcal{C}$ is any category, a diagram $I \to \mathcal{C}$ in which I is a co-filtered category is called a **pro-object** over $\mathcal{C}$. When $\mathcal{C} = \mathcal{T}op$, we will call such an object simply a **pro-space**.

3.5. Rigid covers. Let X be a topological space, and recall the category $\mathrm{OpCov}(X)$ of indexed open coverings of X. Also recall that we constructed a functor $\pi_0\check{C}\colon \mathrm{OpCov}(X) \to s\mathcal{S}et$.

Unfortunately the category $\mathrm{OpCov}(X)$ is not co-filtered. It satisfies the dual condition to Definition 3.2(a), since if $\{U_\alpha : \alpha \in A\}$ and $\{V_\beta : \beta \in B\}$ are two indexed open covers then $\{U_\alpha \cap V_\beta : (\alpha, \beta) \in A \times B\}$ is an open cover which refines both of them. But $\mathrm{OpCov}(X)$ does *not* satisfy the dual condition to 3.2(b):

Exercise 3.6. Let $X = [0, 1]$, and consider the open cover $U_1 = [0, \frac{2}{3})$ and $U_2 = (\frac{1}{3}, 1]$. Let $\mathcal{V}$ be the open cover with $V_1 = X$ and $V_2 = [0, \frac{4}{5})$. We can produce one refinement $f\colon \mathcal{U} \to \mathcal{V}$ by mapping both U_1 and U_2 to V_1. We can produce another refinment $g\colon \mathcal{U} \to \mathcal{V}$ by mapping U_1 to V_2 and U_1 to V_1. Check that there is no cover $\mathcal{W}$ refining $\mathcal{U}$ such that the two maps $f, g\colon \mathcal{W} \to \mathcal{V}$ are the same.

There is a modification of $\mathrm{OpCov}(X)$, suggested originally by Lubkin, which *is* cofiltered. This brings us to the notion of a *rigid* open cover: this is an ordinary open cover such that for each point $x \in X$ we have chosen a distinguished open set of the cover containing it. We can describe things more formally as follows:

Definition 3.7. *A* ***rigid open cover*** *of a topological space X is an indexed open cover $\{U_\alpha : \alpha \in A\}$ together with a choice, for every point $x \in X$, of an index α_x such that $x \in U_{\alpha_x}$.*

Alternatively, we can say that a rigid open cover of X is an indexed open cover together with a non-continuous section of the map $\coprod_{\alpha \in A} U_\alpha \to X$ (that is, a section in the category of sets rather than topological spaces).

Let $\{U_\alpha : \alpha \in A\}$ and $\{V_\beta : \beta \in B\}$ be two rigid covers of X. A **map of rigid open covers** $\mathcal{U} \to \mathcal{V}$ is a map of spaces $\coprod U_\alpha \to \coprod V_\beta$ which makes the following two diagrams commute:

$$\begin{tikzcd} \coprod U_\alpha \arrow[rr] \arrow[dr] & & \coprod V_\beta \arrow[dl] \\ & X & \end{tikzcd} \qquad \begin{tikzcd} \coprod U_\alpha \arrow[rr] & & \coprod V_\beta \\ & X \arrow[ul] \arrow[ur] & \end{tikzcd}$$

(the first should be considered as a diagram of topological spaces, the second only as a diagram of sets). More concretely, to give a map of rigid open covers $\mathcal{U} \to \mathcal{V}$ means to give a function $f\colon A \to B$ such that for each $\alpha \in A$ the open set U_α is contained in $V_{f(\alpha)}$, and such that for each $x \in X$ one has $f(\alpha_x) = \beta_x$.

Definition 3.8. *Let $\{U_\alpha : \alpha \in A\}$ be a rigid open cover. We will say that this is **ultra rigid** if every $\alpha \in A$ is equal to α_x for some $x \in X$.*

Exercise 3.9. Every rigid open cover can be refined by an ultra-rigid open cover. If $\mathcal{U}$ is ultra-rigid and $\mathcal{V}$ is any rigid open cover, then there is at most one map $\mathcal{U} \to \mathcal{V}$.

Proposition 3.10. *Let X be a topological space. The category* $\mathrm{RgdOpCov}(X)$ *of rigid open covers of X is cofiltered.*

Proof. Left as an exercise. □

Remark 3.11. This business with rigid open covers is sort of a "cheap trick" for getting us a cofiltered indexing category. Rather than consider $\pi_0\check{C}\colon \mathrm{OpCov}(X) \to s\mathcal{S}et$, we can work with $\pi_0\check{C}\colon \mathrm{RgdOpCov}(X) \to s\mathcal{S}et$. But why do we need to do this?

As alluded to in Section 1, one thing we will do with systems $D\colon I \to s\mathcal{S}et$ is to define their cohomology. For instance, we will define $H^n(D) = \mathrm{colim}_i\, H^n(D_i)$. Colimits like this behave best, and are most easily computed, when the indexing category I^{op} is filtered—or equivalently, when I is cofiltered. This is perhaps only a minor convenience when dealing with singular cohomology, but we will see in Section 6 that when one starts to work with generalized cohomology theories the cofiltered hypothesis is absolutely necessary for things to work nicely.

There is another way to get a cofiltered indexing category out of $\mathrm{OpCov}(X)$. Observe that if one has two different refinements of open covers $\alpha, \beta\colon \mathcal{U} \to \mathcal{V}$, then the two induced maps on Čech nerves $\check{C}(\mathcal{U}) \to \check{C}(\mathcal{V})$ are simplicially homotopic. So the two maps $\pi_0\check{C}(\mathcal{U}) \to \pi_0\check{C}(\mathcal{V})$ are also simplicially homotopic, and hence they induce *the same* map on singular cohomology. This says that the diagram $H^n(\pi_0\check{C})\colon \mathrm{OpCov}(X) \to \mathcal{A}b$ factors through the category $\mathrm{OpCov}_h(X)$ in which we have identified all maps with the same domain and codomain. The category $\mathrm{OpCov}_h(X)$ is easily seen to be cofiltered. If $D\colon \mathrm{OpCov}(X) \to s\mathcal{S}et$ is a system having the property that any two refinements induce simplicially homotopic maps, we could define

$$H^n(D) = \operatorname*{colim}_{i \in \mathrm{OpCov}_h(X)^{op}} H^n(D_i).$$

This is another way of getting ourselves a filtered colimit to work with. But this approach does not work for generalized cohomology theories, for reasons we will discuss more in Section 6.

3.12. Generalized Čech complexes. It is very useful to realize that the Čech complex of an open cover can be generalized, so that one gets a Čech complex for any *map*. Specifically, for any map $f\colon E \to X$ the associated Čech complex $\check{\boldsymbol{C}}\boldsymbol{(f)}$ is the simplicial space

$$[n] \mapsto E \times_X E \times_X \cdots \times_X E \quad (n+1 \text{ factors}).$$

So the nth level is the space of all tuples $(e_0, \dots, e_{n+1})$ such that all the e_i's map to the same point in X. The face and degeneracy operators correspond to omitting and repeating entries, as usual.

The map $f\colon E \to X$ gives an augmentation $\check{C}(f) \to X$, and so we again get a map $\operatorname{hocolim} \check{C}(f) \to X$.

If $\{U_\alpha\}$ is an open cover of X, then the Čech complex for the map

$$\coprod_{\alpha \in A} U_\alpha \to X$$

is precisely the simplicial space $\check{C}(\mathcal{U})$ defined in Section 2.

THEOREM 3.13. *Let $f\colon E \to X$ be locally split, in the sense that each point $x \in X$ has an open neighborhood U such that $f|_{f^{-1}(U)}\colon f^{-1}(U) \to U$ has a splitting. Then* $\operatorname{hocolim} \check{C}(f) \to X$ *is a weak equivalence.*

The above result is a special case of [**DI**, Prop. 4.10].

EXAMPLE 3.14. Let G be a discrete topological group, and suppose that we have a free G-space E which is contractible. Let $B = E/G$, and let $\pi\colon E \to B$ be the quotient map. This is a covering space, and so is certainly locally split. A point in $E \times_B E \times_B \cdots \times_B E$ is a tuple $(e_0, \dots, e_n)$ such that $\pi(e_0) = \pi(e_j)$ for all j. But since E has a free G-action and π is just the quotient map $E \to E/G$, there exist unique elements $g_i \in G$ such that $e_i = g_i e_{i+1}$. Then we have

$$(e_0, \dots, e_n) = (g_0 g_1 \dots g_{n-1} e_n\,,\, g_1 g_2 \dots g_{n-1} e_n\,,\, \dots\,,\, g_{n-1} e_n\,,\, e_n).$$

Using this isomorphism $E \times_B E \times_B \cdots \times_B E \cong G^n \times E$, i.e. the one given by

$$(e_0, \dots, e_n) \mapsto (g_0, \dots, g_{n-1}, e_n),$$

we find that $\check{C}(\pi)$ is isomorphic as a simplicial space to the usual two-sided bar construction $B(*, G, E)$. Theorem 3.13 therefore tells us that $|B(*, G, E)| \simeq B$.

Now, the projection $E \to *$ gives us a map of simplicial spaces $B(*, G, E) \to B(*, G, *)$ which is an objectwise weak equivalence since E is contractible. So $|B(*, G, *)| \simeq B$, and in this way we recover the usual bar construction for the space BG.

EXAMPLE 3.15. Just as in the previous example, if G is a discrete group and $p\colon E \to B$ is a principal G-bundle then the simplicial space $\check{C}(p)$ is isomorphic to the two-sided bar construction $B(*, G, E)$. So Theorem 3.13 shows that $|B(*, G, E)| \simeq B$. If $\mathcal{E}G$ denotes the groupoid with one object and endomorphism group G, then the G-action on E map be interpreted as giving a functor $\mathcal{E}G \to \mathcal{T}op$ (sending the unique object to E). The simplicial replacement of this diagram is just $B(*, G, E)$, and so $|B(*, G, E)|$ is a model for the homotopy orbit space E_{hG}. Thus we find that $E_{hG} \simeq B$.

Remark 3.16. Theorem 3.13 is certainly not the most general result one can prove along these lines. In fact, it is almost true that $\operatorname{hocolim}\check{C}(f) \to X$ is a weak equivalence whenever f is surjective. There are counterexamples, but they are somewhat exotic. I do not know a nice description of all the maps for which the result holds, and the locally split case will suffice for our purposes in this chapter. In fact, we mostly care about the case where f is a covering space, although we do need something slightly more general:

Definition 3.17. *An* ***étale cover*** *of a topological space X consists of an open cover $\{U_i\}$ together with a covering space $f_i\colon E_i \to U_i$ for each i.*

Often we will consider the associated map $f\colon \coprod_i E_i \to X$. An ***étale covering map*** *is any map of this form. Note that such maps are locally split.*

One can define a map of étale covers in an analagous way to how we defined maps of open covers. It is perhaps easier to phrase things in terms of étale covering maps, in which case a map from $E_1 \to X$ to $E_2 \to X$ is simply a map $E_1 \to E_2$ making the evident triangle commute.

Let us introduce the following categories. $\mathrm{OpCov}(X)$ is the category whose objects are the (indexed) open covers of X, where the maps are refinements. $\mathrm{CovSp}(X)$ is the category of covering spaces of X, where maps are just maps of covering spaces. And finally, $\mathrm{EtCov}(X)$ is the category of étale covering maps, where the morphisms are as defined above. Note that there are inclusions

$$\mathrm{OpCov}(X) \hookrightarrow \mathrm{EtCov}(X) \hookleftarrow \mathrm{CovSp}(X)$$

and that these are inclusions of full subcategories.

Consider the functor $\pi_0\check{C}\colon \mathrm{EtCov}(X) \to s\mathcal{S}et$, sending $p\colon E \to X$ to $\pi_0\check{C}(p)$. We think of this diagram as another system of combinatorial approximations to X. It generalizes the previous system obtained from Čech complexes of open covers, in the sense that the latter can be obtained by restricting our étale system to the subcategory $\mathrm{OpCov}(X) \hookrightarrow \mathrm{EtCov}(X)$.

3.18. Exercises. We close this section with some exercises, the results of which will be needed later.

Exercise 3.19. Define a **rigid étale cover** of a topological space X to be an ordinary étale cover $E \to X$ together with a non-continuous section $X \to E$ (that is, a section in the category of sets). A map of rigid étale covers is just a map of étale covers which is compatible with the sections.

Prove that the category $\mathrm{RgdEtCov}(X)$ of rigid étale covers of X is cofiltered. [First define the notion of an ultra-rigid étale cover. Prove that every rigid étale cover can be refined by an ultra-rigid cover, and that if $\mathcal{U}$ is an ultra-rigid cover and $\mathcal{V}$ is a rigid étale cover then there is at most one map $\mathcal{U} \to \mathcal{V}$.]

Exercise 3.20 (Cofinality).

(a) A functor $\alpha\colon I \to J$ between filtered categories is called **cofinal** if the following two conditions are satisfied:
  (1) For every $j \in J$, there exists an $i \in I$ and a map $\alpha(i) \to j$.
  (2) For every $j \in J$, $i \in I$, and two maps $j \rightrightarrows \alpha(i)$, there exists a map $i \to i'$ such that the two composites $j \rightrightarrows \alpha(a) \to \alpha(i')$ are equal.

  Under these hypotheses, prove that if $D\colon J \to \mathcal{A}b$ is a functor then the natural map $\operatorname{colim}_I(D\alpha) \to \operatorname{colim}_J D$ is an isomorphism.

(b) A functor $\alpha\colon I \to J$ between cofiltered categories is called **final** if the functor $\alpha^{op}\colon I^{op} \to J^{op}$ is cofinal. Prove that the functor $\mathrm{RgdOpCov}(X) \to \mathrm{RgdEtCov}(X)$ is final.

REMARK 3.21. If $X\colon J \to s\mathcal{S}et$ is a pro-space and $\alpha\colon I \to J$ is final, then the pro-spaces X and $X\alpha$ behave the same "in the limit". So the point of the above exercise is that the information in the pro-spaces $\pi_0\check{C}\colon \mathrm{RgdOpCov}(X) \to s\mathcal{S}et$ and $\pi_0\check{C}\colon \mathrm{RgdOpCov}(X) \to s\mathcal{S}et$ is really the same.

The following exercises concern manipulations with Čech complexes.

EXERCISE 3.22. Let $p\colon E \to Y$ and $f\colon X \to Y$ be any maps. Let $E' = X \times_Y E$, and $p'\colon E' \to X$ be the evident projection. There is a map of simplicial spaces $\check{C}(p') \to \check{C}(p)$.

(a) Prove that $\check{C}(p')$ is isomorphic to the simplicial space obtained by applying the functor $X \times_Y (-)$ to $\check{C}(p)$.
(b) Now assume that $E \to Y$ is a fibration, X and Y are path connected, and that $X \to Y$ is surjective on π_1. Prove that $\pi_0\check{C}(p') \to \pi_0\check{C}(p)$ is an isomorphism of simplicial sets.

EXERCISE 3.23. Let G be a group and $H \subseteq G$ a normal subgroup. The covering space $EG \times_G (G/H) \to BG$ is a principal G/H-bundle, and is therefore classified by a map $BG \to B(G/H)$. Use the previous exercise to show that

$$\check{C}\Big(EG \times_G (G/H) \to BG\Big) \cong B(*, G/H, *).$$

EXERCISE 3.24. Let G be a discrete group. For any left G-set S, let $\pi_0^G(S)$ denote the set of G-orbits G/S (the notation comes from thinking of the elements of an orbit as being 'connected'). If S is a left G-set then the Čech complex $\check{C}(S \to *)$ is a simplicial G-set; we'll denote this just by $\check{C}(S)$, as usual.

(a) Prove that $\pi_0(E) \cong \pi_0^G(S)$, and then use this to show that the two simplicial sets $\pi_0\check{C}(EG \times_G S \to BG)$ and $\pi_0^G(\check{C}(S))$ are isomorphic.
(b) Let $\mathbb{Z}[\check{C}(S)]$ be the associated chain complex of $\mathbb{Z}[G]$-modules obtained by taking the alternating sum of the face maps. Prove that $\mathbb{Z}[\check{C}(S)]$ is acyclic, and $H_0(\mathbb{Z}[\check{C}(S)]) \cong \mathbb{Z}$ as a $\mathbb{Z}[G]$-module.
(c) Show that $H^n(\check{C}(EG \times_G S \to BG))$ is the nth cohomology of the cochain complex $\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[\check{C}(S)], \mathbb{Z})$.
(d) If $S = G/H$ where H is a normal subgroup, prove that $S \times S$ (regarded as a left G-set) is a disjoint union of copies of G/H. Conclude that $\mathbb{Z}[\check{C}(S)]$ is a free resolution of $\mathbb{Z}$ by free $\mathbb{Z}[G/H]$-modules, and deduce that $H^*(\check{C}(EG \times_G S)) \cong H^*(G/H; \mathbb{Z})$. Compare with the result of Exercise 3.23.

4. Cohomology via étale coverings

In this section we return to algebraic geometry and continue our attempt to attach homotopy invariants to algebraic varieties.

4.1. A motivating example. We begin with a basic example which demonstrates most of the important points. We will treat this example in some detail.

Let our ground field be $k = \mathbb{C}$ and let $X = \mathbb{A}^1 - 0$. For any $n \geq 0$, the map $\rho_n \colon \mathbb{A}^1 - 0 \to \mathbb{A}^1 - 0$ given by $z \mapsto z^n$ is a covering space. Let $E_n = \mathbb{A}^1 - 0$, so that we can write ρ_n as a map $E_n \to X$.

Consider the simplicial scheme $\check{C}(\rho_n)$ given by $[k] \mapsto E_n \times_X \cdots \times_X E_n$ ($k+1$ factors). Let $\pi_0\check{C}(\rho_n)$ denote the simplicial set obtained by replacing the scheme in each dimension by the set of connected components of its underlying topological space. Our aim is to investigate these simplicial sets and see whether they give reasonable approximations to the homotopy type of X (which is a circle).

First, we claim that $E_n \times_X E_n$ is a disjoint union of n copies of E_n. To see this, note that $E_n \to X$ is spec of the map of rings $k[z, z^{-1}] \to k[z^{1/n}, z^{-1/n}]$. That is, if we let $R = k[z, z^{-1}]$ then we are looking at $R \to R[x]/(x^n - z)$. Then $E_n \times_X E_n$ is spec of the ring

$$R[x]/(x^n - z) \otimes_R R[x]/(x^n - z) \cong k[z^{1/n}, z^{-1/n}, x]/(x^n - z).$$

Let ζ be a primitive nth root of unity in k. Then in $k[z^{1/n}, z^{-1/n}, x]$, the polynomial $x^n - z$ splits as

$$x^n - z = (x - z^{1/n})(x - \zeta z^{1/n})(x - \zeta^2 z^{1/n}) \cdots (x - \zeta^{n-1} z^{1/n}).$$

So the coordinate ring of $E_n \times_X E_n$ splits as

$$k[z^{1/n}, z^{-1/n}] \times k[z^{1/n}, z^{-1/n}] \times \cdots \times k[z^{1/n}, z^{-1/n}]$$

(n factors). This says precisely that $E_n \times_X E_n$ is isomorphic to a disjoint union of n copies of E_n. One then finds that $E_n \times_X E_n \times_X E_n$ is a disjoint union of n^2 copies of E_n, and so on.

So in dimension k, the simplicial scheme $\check{C}(\rho_n)$ contains a disjoint union of n^k copies of E_n. Either by brute force inspection or by comparison with Exercise 3.23 (in the case $G = \mathbb{Z}$ and $H = (n)$), one now sees that

$$\pi_0\check{C}(\rho_n) \cong B\mathbb{Z}/n.$$

Is this a space a reasonable 'approximation' to S^1? The homotopy groups are somewhat similar, but the cohomology groups are very different. Let's recall how to compute the latter.

Let A be an abelian group. The cohomology groups $H^*(B\mathbb{Z}/n; A)$ are isomorphic to the groups $\mathrm{Ext}^*_{\mathbb{Z}[\mathbb{Z}/n]}(\mathbb{Z}, A)$ where $\mathbb{Z}[\mathbb{Z}/n]$ is the group ring and both $\mathbb{Z}$ and A have the trivial module structure. Let $G = \mathbb{Z}/n$ and write g for some choice of generator. Then the group ring $\mathbb{Z}[G]$ is the ring $\mathbb{Z}[g]/(g^n - 1)$. The module $\mathbb{Z} = \mathbb{Z}[G]/(g-1)$ has free resolution given by

$$\cdots \longrightarrow \mathbb{Z}[G] \xrightarrow{N} \mathbb{Z}[G] \xrightarrow{1-g} \mathbb{Z}[G] \xrightarrow{N} \mathbb{Z}[G] \xrightarrow{1-g} \mathbb{Z}[G] \longrightarrow \mathbb{Z}$$

where $N = 1 + g + g^2 + \cdots + g^{n-1}$. It follows that the Ext groups we want are the cohomology of the complex

$$\cdots \xleftarrow{0} A \xleftarrow{n} A \xleftarrow{0} A.$$

So in particular,

$$H^*(B\mathbb{Z}/n;\mathbb{Z}) \cong \begin{cases} \mathbb{Z} & \text{if } * = 0 \\ 0 & \text{if } * > 0 \text{ is odd} \\ \mathbb{Z}/n & \text{if } * > 0 \text{ is even,} \end{cases}$$

and

$$H^*(B\mathbb{Z}/n;\mathbb{Z}/k) \cong \begin{cases} \mathbb{Z}/k & \text{if } * = 0 \\ \mathbb{Z}/(n,k) & \text{otherwise.} \end{cases}$$

So in some sense, this is bad news. The cohomology groups of $\check{C}(\rho_n)$ don't look anything like the cohomology groups of S^1. The way out of this is to not look at $\check{C}(\rho_n)$ just on its own like this. We need to look at these spaces for all possible values of n, and look at them all together.

Consider the two covering spaces $E_n \to X$ and $E_k \to X$. We know that there will be a map $E_n \to E_k$ making the evident triangle commute if and only if k divides n, in which case there are k such maps. If we require in addition that $E_n \to E_k$ send 1 to 1, then there is only one such map: the map $\mathbb{A}^1 - 0 \to \mathbb{A}^1 - 0$ given by $z \mapsto z^{n/k}$.

Let I be the poset of non-negative integers, where n is less than k if and only if n divides k. Regard I as a category in the usual way. The we have a functor $I \to \mathcal{S}ch/X$ sending n to $E_n \to X$, and if n divides k then the map $n \to k$ is sent to the unique map of covering spaces $E_n \to E_k$ sending 1 to 1. By applying the Čech construction to each $E_n \to X$, we obtain a functor $\pi\check{C}E\colon I \to s\mathcal{S}et$ given by $n \mapsto \pi_0\check{C}(\rho_n)$.

We would like to think of $\pi_0\check{C}E$ as a 'system of approximations' to the homotopy type of X. Any individual object $\pi\check{C}E_n$ is not a particularly good approximation, but maybe things are better if we take them altogether. Returning to the issue of cohomology, we now need to compute the maps $H^*(\pi\check{C}E_k; A) \to H^*(\pi\check{C}E_n; A)$ when k divides n. Note that this is the map $H^*(B\mathbb{Z}/k; A) \to H^*(B\mathbb{Z}/n; A)$.

Exercise 4.2. Let $\mathbb{Z}/nr \to \mathbb{Z}/n$ be the usual projection. We need to compute the map $H^*(B\mathbb{Z}/n; A) \to H^*(B\mathbb{Z}/nr; A)$. Write $G = \mathbb{Z}/nr$ and $H = \mathbb{Z}/n$. So we are interested in the map of Ext-groups

$$\phi\colon\ \mathrm{Ext}^*_{\mathbb{Z}[H]}(\mathbb{Z}, A) \to \mathrm{Ext}^*_{\mathbb{Z}[G]}(\mathbb{Z}, A)$$

induced by the map of rings $\pi\colon \mathbb{Z}[G] \to \mathbb{Z}[H]$. Note that this is a quotient map, and $\mathbb{Z}[H]$ may be identified with $\mathbb{Z}[G]/(g^n - 1)$.

To compute the map ϕ we construct free resolutions for $\mathbb{Z}$ over $\mathbb{Z}[G]$ and $\mathbb{Z}[H]$ and then get a comparison map of resolutions:

$$\begin{tikzcd}
\cdots \arrow[r] & \mathbb{Z}[G] \arrow[r, "N_{nr}"] \arrow[d, "\phi_4"] & \mathbb{Z}[G] \arrow[r, "1-g"] \arrow[d, "\phi_3"] & \mathbb{Z}[G] \arrow[r, "N_{nr}"] \arrow[d, "\phi_2"] & \mathbb{Z}[G] \arrow[r, "1-g"] \arrow[d, "\phi_1"] & \mathbb{Z}[G] \arrow[r] \arrow[d, "\phi_0"] & \mathbb{Z} \arrow[d, equal] \\
\cdots \arrow[r] & \mathbb{Z}[H] \arrow[r, "N_n"] & \mathbb{Z}[H] \arrow[r, "1-g"] & \mathbb{Z}[H] \arrow[r, "N_n"] & \mathbb{Z}[H] \arrow[r, "1-g"] & \mathbb{Z}[H] \arrow[r] & \mathbb{Z}
\end{tikzcd}$$

Here $N_{nr} = 1 + g + g^2 + \cdots + g^{nr-1}$, $N_n = 1 + g + g^2 + \cdots + g^{n-1}$, and everything in the diagram is a map of $\mathbb{Z}[G]$-modules.

It is easy to see that we may take $\phi_0 = \phi_1 = \pi$. The map ϕ_2 must sent 1 to $1 + g^n + g^{2n} + \cdots + g^{(r-1)n}$, and we may take ϕ_3 to be the same map. Proceeding inductively, we find

$$\phi_4 = \phi_5 = \left[1 + g^n + g^{2n} + \cdots + g^{(r-1)n}\right]^2,$$

and so on.

Finally, we have that for each $k \geq 0$ the induced maps $\phi^{2k}\colon \operatorname{Ext}^{2k}_{\mathbb{Z}[G]}(\mathbb{Z}, A) \to \operatorname{Ext}^{2k}_{\mathbb{Z}[H]}(\mathbb{Z}, A)$ and $\phi^{2k+1}\colon \operatorname{Ext}^{2k+1}_{\mathbb{Z}[G]}(\mathbb{Z}, A) \to \operatorname{Ext}^{2k+1}_{\mathbb{Z}[H]}(\mathbb{Z}, A)$ are both multiplication by r^k. The exercise is to check all the details here.

Finally, consider the groups $\operatorname{colim}_{I^{op}} H^*(\pi\check{C}E_n; A)$ for various abelian groups A. Here are some things we can conclude, based on the computations from the above exercise:

$$\operatorname*{colim}_{I^{op}} H^*(\pi\check{C}E_n; \mathbb{Z}/l^e) \cong \begin{cases} \mathbb{Z}/l^e & \text{if } * \in \{0,1\} \\ 0 & \text{otherwise.} \end{cases}$$

Also,

$$\operatorname*{colim}_{I^{op}} H^*(\pi\check{C}E_n; \mathbb{Z}) \cong \begin{cases} \mathbb{Z} & \text{if } * \in \{0,1\}, \\ \mathbb{Q}/\mathbb{Z} & \text{if } * = 2, \\ 0 & \text{otherwise.} \end{cases}$$

With integral coefficients, these colimit groups still do not look very much like the cohomology of S^1. But with finite coefficients things look right!

Exercise 4.3. Prove that if A is any abelian group then

$$\operatorname*{colim}_{I^{op}} H^*(\pi\check{C}E_n; A) \cong \begin{cases} A & \text{if } * = 0 \\ tors(A) \cong \operatorname{Tor}_1(A, \mathbb{Q}/\mathbb{Z}) & \text{if } * = 1 \\ A \otimes \mathbb{Q}/\mathbb{Z} & \text{if } * = 2 \\ 0 & \text{otherwise.} \end{cases}$$

Conclude that $\operatorname{colim}_{I^{op}} H^*(\pi\check{C}E_n; A)$ is isomorphic to $H^*(S^1; A)$ whenever A is a torsion group.

4.4. First attempts at generalizations. Start with a variety X over $\mathbb{C}$. Choose a point $x \in X(\mathbb{C})$, and consider the category $I_{(X,x)}$ whose objects are pairs $(E \to X, e)$ where $E \to X$ is a map of varieties which is topologically a covering space, and $e \in E(\mathbb{C})$ is in the preimage of x. A map $(E \to X, e) \to (E' \to X, e')$ is a map of varieties over X sending e to e'. It is easy to see that there is at most one map between any two objects.

We obtain a functor $\pi\check{C}E\colon I_{(X,x)} \to s\mathcal{S}et$ sending $E \to X$ to $\pi_0\check{C}(E)$, and we consider the groups

$$\operatorname*{colim}_{I^{op}_{(X,x)}} H^*(\pi\check{C}(E_i); A) \tag{4.5}$$

for various abelian groups A. In the case $X = \mathbb{A}^1 - 0$, we found that these agreed with the groups $H^*(X; A)$ when A is a torsion group. Does this work for other varieties?

A moment's thought shows that it cannot possibly work for most varieties. For instance, take $X = \mathbb{C}P^1$. Then because X is simply connected, it does not have any nontrivial covering spaces at all! The groups in (4.5) are therefore zero, whereas $X \simeq S^2$ and has a non-vanishing H^2. In fact, it is not hard to convince oneself that this naive procedure cannot work for any variety which is not topologically a $K(G,1)$.

To get a better procedure what we must do is mix Zariski covers with the covering space approach. For instance, consider $X = \mathbb{C}P^1$. Take the standard Zariski cover $\{U_0, U_1\}$ where $U_0 \cong U_1 \cong \mathbb{A}^1$ and $U_0 \cap U_1 \cong \mathbb{A}^1 - 0$. Each of U_0, U_1, and $U_0 \cap U_1$ is topologically a $K(G,1)$ (for the first two G is even trivial), and so we can hope to reconstruct their cohomology groups by looking at algebraic covering spaces. At the same time, we know that topologically X is the homotopy colimit of

$$U_0 \leftarrow U_0 \cap U_1 \rightarrow U_1$$

and so we can reconstruct the cohomology groups of X from those of U_0, U_1, and $U_0 \cap U_1$ via the Mayer-Vietoris sequence.

The above procedure is a bit clunky at the moment, but eventually we'll develop a slick way of organizing everything into one package. But to summarize, here are the main points we have discussed so far:

(1) Given an algebraic variety X over $\mathbb{C}$, choose (if we can) a Zariski cover $\{U_\alpha\}$ so that all the iterated intersections are 'good', in the sense that our construction from (4.5) gives their correct cohomology groups with torsion coefficients. This will *at least* require that the iterated intersections are topologically $K(G,1)$'s.
(2) For each $U_\sigma = U_{\sigma_1} \cap \cdots \cap U_{\sigma_k}$, consider the Čech nerves for all the algebraic covering spaces of U_σ and take the colimit of their cohomology groups.
(3) Use the cohomology groups in (2) to reconstruct—or approximate—the cohomology of X via the Čech complex of the $\{U_\alpha\}$ cover.

Čech complexes appear twice in the above procedure, first for the covering spaces of the U_σ's and then for the Zariski cover $\{U_\alpha\}$. Some of the clunkiness of the above description will be removed by putting Zariski covers and covering spaces together into one notion—the so-called *étale covers*. Then the procedure will only have one set of Čech complexes, and one colimit. This is only a matter of bookeeping.

A more serious matter is the question of whether we can really choose 'small enough' Zariski open sets U_α so that the construction from (4.5) gives the correct cohomology groups. We will investigate this next.

4.6. Artin neighborhoods. We start by observing that varieties X which have the homotopy type of a $K(G,1)$ are quite plentiful. Every smooth, projective algebraic curve is topologically a genus g torus, and such things are all $K(G,1)$'s. If the curve is smooth and non-projective then it came from a smooth projective curve by removing finitely-many points—and such things have the homotopy type of a wedge of circles. So every smooth algebraic curve is a $K(G,1)$. In fact, below we will prove the following result:

THEOREM 4.7. *Let X be a complex algebraic variety, $x \in X$ be a smooth point, and let U be a Zariski open set containing x. Then there is another Zariski open set V satisfying $x \in V \subseteq U$ and such that V is topologically a $K(G,1)$. Moreover, by choosing V appropriately one can arrange that G is freely constructible in the sense of the following definition.*

DEFINITION 4.8. *A group will be called* ***freely constructible*** *if it belongs to the smallest class of groups $\mathcal{S}$ satisfying the following two properties:*

(i) $\mathcal{S}$ *contains the trivial group;*
(ii) *If* $1 \to F \to G \to Q \to 1$ *is an exact sequence where* Q *belongs to* $\mathcal{S}$ *and* F *is a finitely-generated free group, then* G *belongs to* $\mathcal{S}$.

A complete proof of the above theorem will be given later in the exercises. For now we will be content with a vague sketch, to demonstrate the basic ideas:

SKETCH OF PROOF OF THEOREM 4.7. The proof is by induction on the dimension of X. The dimension zero case is trivial. For the dimension one case, U is topologically a genus g torus minus a finite set of points, and therefore it is a $K(G,1)$. By removing at least one point from the g-torus, we can ensure that G is free (and hence freely constructible).

Now suppose that the dimension of X is n, where $n \geq 2$. By replacing X with a Zariski neighborhood of x, we can assume X is affine. We can then embed X in some $\mathbb{C}P^N$ and take the closure, so we may in fact assume X is a closed subvariety of $\mathbb{C}P^N$.

A linear map $f\colon \mathbb{C}^{N+1} \to \mathbb{C}^n$ induces a map $F\colon \mathbb{C}P^N - \mathbb{P}(\ker f) \to \mathbb{C}P^n$ by sending $[\mathbf{x}]$ to $[f(\mathbf{x})]$. A generically chosen f will be a surjection, and each fiber of F will be a copy of $\mathbb{C}P^{N-n+1}$.

Let π denote the composite

$$X - (X \cap \mathbb{P}(\ker f)) \hookrightarrow \mathbb{C}P^N - \mathbb{P}(\ker f) \to \mathbb{C}P^n.$$

Again for generically chosen f, $\mathbb{P}(\ker f)$ will be a copy of $\mathbb{C}P^{N-n}$. A generic $\mathbb{C}P^{N-n}$ in $\mathbb{C}P^N$ will meet X in exactly d points, where d is the degree of X. By choosing the linear map generically we can assume that x is not one of these d points, and that x is not a critical point of π.

Let X_{sm} be the open subvariety of X consisting of the smooth points. Let $X' = X_{sm} - (X_{sm} \cap \mathbb{P}(\ker f))$, and $U' = U \cap X'$. The fibers of $\pi|_{X'}\colon X' \to \mathbb{C}P^n$ are generically one-dimensional (they are obtained by intersecting X' with the $\mathbb{C}P^{N-n+1}$'s forming the fibers of F). So locally around $\pi(x)$ the fibers look like a genus g torus minus a finite number of points. We can then choose a Zariski neighborhood $\pi(x) \in J \subseteq \mathbb{C}P^n$ such that the number of points being removed from the torus is the same in all fibers. By induction, there exists a Zariski neighborhood $J' \subseteq J$ of $\pi(x)$ that is topologically a $K(G,1)$, with G freely constructible. Let $U'' = \pi^{-1}(J')$.

We have arranged things so that $\pi\colon U'' \to J'$ is a fibration where the fibers are $K(H,1)$'s, with H free. As J' is a $K(G,1)$, it follows from the long exact homotopy sequence that U'' is a $K(G',1)$ where G' sits in the short exact sequence $1 \to H \to G' \to G \to 1$. Since G is freely constructible, so is G', and this completes the proof. $\square$

The first moral of the theorem is that varieties which are $K(G,1)$'s are very common, which is good news. Assuming X is such a variety, we will now investigate whether the cohomology groups can be reconstructed by the method of (4.5). We need to understand how the algebraic coverings spaces of X compare to the topological covering spaces. The case $\mathbb{A}^1 - 0$ turns out to be somewhat typical. There we found that every finite covering space of X could be realized as an algebraic variety, but that the infinite covering space could not. Here is the general theorem:

THEOREM 4.9 (Riemann existence theorem). *Let* X *be a complex algebraic variety. If* $E \to X$ *is a map of algebraic varieties which is topologically a covering* *space, then the fibers are finite. Conversely, any finite covering space of X can be realized by a map of algebraic varieties.*

This was proven by Riemann in the case where X is a smooth projective curve. The general case follows from results by Grauert and Remmert [**GR**], with the aid of Serre's GAGA [**Se5**]. We will not recount the proof here, as it is very technical; but see [**SGA4**, XI,Theorem 4.3].

Here is a slightly better version of the theorem, also from [**SGA4**]:

THEOREM 4.10 (Riemann existence theorem, improved version). *Let X be a complex algebraic variety. Let $\mathrm{Cov}_{alg}(X)$ be the category whose objects are maps of varieties $E \to X$ which topologically are covering spaces; the morphisms are maps of varieties over X. Let $\mathrm{Cov}_{finite}(X)$ be the category of all topological covering spaces of X with finite fibers. Then there is an evident functor $\mathrm{Cov}_{alg}(X) \to \mathrm{Cov}_{finite}(X)$, and this is an equivalence of categories.*

SKETCH OF PROOF. The previous result is the statement that the functor is surjective on isomorphism classes. It remains to show that the functor induces bijections on hom-sets.

It is clear that the induced maps of hom-sets are injective, because a map of complex varieties $Z \to W$ is determined by the induced map $Z(\mathbb{C}) \to W(\mathbb{C})$. Ref????

The proof of surjectivity hinges on the following topological fact. Suppose that $E \to B$ and $F \to B$ are two covering spaces of a connected space B. Then maps of covering spaces $E \to F$ are in bijective correspondence with components of the pullback $E \times_B F$, with each map corresponding to its graph. Keeping this in mind, uppose $E_1 \to X$ and $E_2 \to X$ are maps of varieties which topologically are covering spaces, and suppose $f\colon E_1(\mathbb{C}) \to E_2(\mathbb{C})$ is a covering map. Then f determines a component of the pullback $E_1(\mathbb{C}) \times_{X(\mathbb{C})} E_2(\mathbb{C})$. Let Y denote the corresponding component of the scheme $E_1 \times_X E_2$. The projection $Y \to E_1$ is an isomorphism, so by composing the inverse with the projection $Y \to E_2$ one obtains an algebraic map $E_1 \to E_2$ which is readily checked to induce the map f. □

So our category $I_{(X,x)}$ may be identified with the category of pointed topological covers of finite degree. It will be useful to recall the fundamental theorems of covering space theory.

EXERCISE 4.11. Let Z be a space which is semi-locally simply connected (e.g. a CW-complex). Let $z \in Z$ and let $G = \pi_1(Z, z)$. Then there is an equivalence of categories between the category of covering spaces of Z and the category of left G-sets. The equivalence sends a covering space $p\colon E \to Z$ to the fiber $p^{-1}(z)$, equipped with the monodromy action of G. If $\tilde{Z}$ is the universal covering space of Z then it has an evident right G-action, and the equivalence sends a left G-set S to the covering space $\tilde{Z} \times_G S$.

The equivalence of categories restricts to an equivalence between the full subcategory of connected covering spaces and the full subcategory of transitive G-sets.

Finally, if we consider the category of *pointed*, connected covering spaces (a connected covering space equipped with a choice of point in the pre-image of z) then this is equivalent to the category of pointed, transitive G-sets. This latter category is equivalent to the opposite of the category of subgroups of G, with maps the inclusions, by sending a pointed G-set (S, s) to the stabilizer of s. The functor in the other direction sends the subgroup H to the transitive G-set G/H, pointed by the coset eH (where e is the identity of G).

So when X is a $K(G,1)$, the category of pointed, connected covering spaces is equivalent to the opposite of the category of subgroups of G. The subgroup H corresponds to the covering space $EG \times_G (G/H) \to BG$, and the finite covering spaces correspond to the subgroups of finite index.

Let $S_f(G)$ be the category of subgroups of G of finite index. Let $S_{nf}(G)$ be the category of normal subgroups of finite index. We first remark that $S_{nf}(G)$ is final in $S_f(G)$. This is because if H is any subgroup of finite index then $H_n = \bigcap_{g\in G} gHg^{-1}$ is a finite intersection of subgroups of finite index, and hence also has finite index. But H_n is clearly normal, and so every subgroup of finite index contains a normal subgroup of finite index.

When G is normal, the covering space $EG\times_G(G/H) \to BG$ is a principal G/H-bundle. So by Exercise 3.23 we have that $\pi_0\check{C}(EG \times_G G/H)$ is the simplicial set $B(*, G/H, *)$ (the usual bar construction for G/H). We are therefore now reduced to considering the following question. If A is a torsion abelian group, will the map

$$\operatorname*{colim}_{H\in S_{nf}(G)} H^*(G/H; A) \to H^*(G; A) \tag{4.12}$$

necessarily be an isomorphism?

It is not hard to find examples where this doesn't hold. For instance, take $G = \mathbb{Q}/\mathbb{Z}$. Every map from $\mathbb{Q}/\mathbb{Z}$ into a finite group is the zero map, and so $\mathbb{Q}/\mathbb{Z}$ does not have any nontrivial subgroups of finite index. Thus, the colimit in (4.12) is the zero group. However, the Serre spectral sequence for the fibration $B\mathbb{Z} \to B\mathbb{Q} \to B(\mathbb{Q}/\mathbb{Z})$ shows immediately that $H^2(\mathbb{Q}/\mathbb{Z}; \mathbb{Z}/p) \cong \mathbb{Z}/p$ (using that the cohomology of $B\mathbb{Q}$ with finite coefficients all vanishes).

Another example is the infinite alternating group A_∞. This is a simple group, and so has no nontrivial normal subgroups at all. The colimit group in (4.12) is again zero. But with some trouble one can see that $H^*(A_\infty; \mathbb{Z}/p)$ is nonzero for all primes p (it is related to the cohomology of the infinite symmetric group Σ_∞, which was computed by Nakaoka).

So (4.12) is not always an isomorphism. But it is an isomorphism in some important examples, as we now explain.

Definition 4.13. *A group G will be called* ***good for profinite completion*** *if (4.12) is an isomorphism for all torsion abelian groups A.*

Theorem 4.14. *The following groups are good for profinite completion:*

(a) Finitely-generated free groups.

(b) The fundamental group of a genus g-torus, for any g.

(c) Any group which is freely constructible in the sense of Definition 4.8.

The proof of this result will be sketched in the exercises below.

Exercise 4.15. Let G be a group and $J \subseteq G$ be a subgroup. For each (left) J-module M, define a G-module $\mathrm{Ind}_G^J(M)$ in the following way. As a set, $\mathrm{Ind}_G^J(M)$ consists of all J-equivariant maps $G \to M$; the abelian group structure is given by pointwise addition. Finally, if $g \in G$ and $f \in \mathrm{Ind}_G^J(M)$, let gf be the map $G \to M$ given by $(gf)(u) = f(ug)$.

(a) Check that $\mathrm{Ind}_G^J(M)$ is indeed a left G-module.

(b) Verify that one has adjoint functors

$$U\colon G-\mathcal{M}od \rightleftarrows J-\mathcal{M}od\colon \operatorname{Ind}_G^J$$

where U is the forgetful functor and is the left adjont in the pair.

(c) Deduce natural isomorphisms $\operatorname{Ext}^p_{G-\mathcal{M}od}(\mathbb{Z}, \operatorname{Ind}_G^J(M)) \cong \operatorname{Ext}^p_{J-\mathcal{M}od}(\mathbb{Z}, M)$ for all J-modules M and all $p \geq 0$. That is, deduce the existence of isomorphisms

$$H^p(G, \operatorname{Ind}_G^J(M)) \cong H^p(J, M).$$

(d) Prove the following statements:

(i) If J has finite index in G and M is a finite J-module, then $\operatorname{Ind}_G^J(M)$ is a finite G-module.

(ii) If M is a G-module then the unit of the adjunction $M \to \operatorname{Ind}_G^H(UM)$ is an injection.

(iii) If $H \subseteq G$ is a normal subgroup and M is a G-module on which H acts trivially, then $\operatorname{Ind}_G^H(UM) = \operatorname{Ind}_{G/H}^e(M)$.

(iv) $\operatorname{Ind}_G^e(A)$ is an injective G-module, for any abelian group A.

The following two exercises are based on [**Se3**, Section 2.6 exercises].

Exercise 4.16. Let G be a group.

(a) If M is a finite G-module, the group action may be regarded as a homomorphism $G \to \operatorname{Aut}(M)$. Deduce that M is the restriction of a finite G/K-module for some normal subgroup $K \subseteq G$ of finite index.

(b) Consider the induced map

$$\psi_n\colon \operatorname*{colim}_{J\subseteq G} H^n(G/J; M) \to H^n(G; M)$$

where the subgroups J range over all normal subgroups of G which are contained in K. (Note that the exact choice of K does not matter, as a different choice will lead to a colimit which is canonically isomorphic to the one above.) Consider the following properties of G:

(A_n) For every finite module M, ψ_p is bijective for all $0 \leq p \leq n$ and injective for $p = n+1$.

(B_n) For every finite module M and all $0 \leq p \leq n$, ψ_p is surjective.

(C_n) For every finite module M and every $x \in H^p(G; M)$, $1 \leq p \leq n$, there exists a finite module $M \subseteq M'$ such that the image of x in $H^p(G; M')$ vanishes.

(D_n) For every finite module M and every $x \in H^p(G; M)$, $1 \leq p \leq n$, there exists a subgroup $J \subseteq G$ of finite index such that the image of x in $H^p(J; M)$ vanishes.

Prove that properties $A_n, \ldots, D_n$ are equivalent as follows.
First, argue that $A_n \Rightarrow B_n \Rightarrow D_n \Rightarrow C_n \Rightarrow B_n$ (the latter by induction on p).
Now prove that $C_n \Rightarrow A_n$ by the following method. First, the surjectivity of the ψ_p has already been argued as part of $C_n \Rightarrow B_n$. So assume $x \in H^p(G/J; M)$ is such that the image of x in $H^p(G; M)$ is zero. Consider the exact sequence $0 \to M \to \operatorname{Ind}_G^J(M) \to M' \to 0$. Argue that x is the image of an element $x_1 \in H^{p-1}(G/J; M')$. Let x_2 be the image of x_1 in $H^{p-1}(G; M')$, and argue that x_2 is the image of an $x_3 \in H^{p-1}(G, \operatorname{Ind}_G^J(M))$. Now use property C_n and induction.

(c) Note that C_0 (and D_0) are trivially true. Prove that $A_1, \ldots, D_1$ are also true.

(d) If G satisfies $A_n, \dots, D_n$ and $J \subseteq G$ has finite index, prove that J also satisfies $A_n, \dots, D_n$.

EXERCISE 4.17. In this exercise we prove Theorem 4.14. Let us say that a group satisfies A_∞ if it satisfies A_n for all n.

(a) First observe if F is a free group then $H^p(F; M) = 0$ for all F-modules M and all $p > 1$. So free groups satisfy A_∞ by Exercise 4.16(c).
(b) Let C be the genus g torus. Prove that for any $n \in \mathbb{N}$ there exists a finite covering space $C' \to C$ such that the induced map $H^2(C) \to H^2(C')$ sends a generator to n times a generator. Deduce that for any finite abelian group A and $x \in H^2(C; A)$, there exists a finite covering space $C' \to C$ with the property that x maps to zero in $H^2(C'; A)$. Using that $C = BG$, where $G = \pi_1(C)$, deduce that G is good for profinite completion (this might involve rehashing some of the arguments from Exercise 4.16).
(c) Finally, we tackle the main case of interest and show that any freely constructible group satisfies A_∞. The case of free groups was dealt with in (a), so assume $1 \to N \to E \to G \to 1$ is a short exact sequence where N is a finitely-generated free group and G satisfies A_∞. We will prove that E satisfies D_n, for all n.

   For any E-module M there is a spectral sequence of the form
$$E_2^{p,q} = H^p(G; H^q(F; M)) \Rightarrow H^{p+q}(E; M),$$
and since F is free this spectral sequence is concentrated along the lines $q = 0$ and $q = 1$. Write $H^p(E; M) = F_0 \supseteq F_1 \supseteq F_2 \supseteq \cdots$ for the filtration that the spectral sequence is converging to. Then the form of the spectral sequence shows that $F_k = 0$ for $k \geq 2$, and we have natural isomorphisms
$$\alpha\colon H^p(G; H^0(F; M))/\operatorname{im} d_2 \xrightarrow{\cong} F_0/F_1$$
and
$$\beta\colon \ker d_2|_{H^{p-1}(G; H^1(F; M))} \xrightarrow{\cong} F_1.$$
Let $x \in H^p(E; M)$ where $p \geq 2$. Choose a class $y_1 \in H^p(G; H^0(F; M))$ which maps to (the coset of) x under α. Since G satisfies A_∞, there is a finite index subgroup $G_0 \subseteq G$ such that the image of y_1 in $H^p(G_0; H^0(F; M))$ vanishes. Let E_0 be the preimage of G_0 under $E \to G$, and consider the sequence $1 \to F \to E_0 \to G_0 \to 1$. Using the naturality of the spectral sequence, we find that the image of x in $H^p(E_0; M)$ lives in F_1. Let $y_2 \in H^{p-1}(G_0; H^1(F; M))$ be a class which maps, under β, to this image of x. Since G_0 satisfies A_∞, there is a finite index subgroup $G_1 \subseteq G_0$ such that y_2 maps to zero in $H^{p-1}(G_0; H^1(F; M))$. A little work completes the argument.

5. Étale maps in algebraic geometry

There are several different ways of saying what an étale map is, all of which are equivalent. We start with the ones which are most easily checked in practice.

Let k be a field and let R be a k-algebra. A **standard étale map** is a map of k-algebras of the form $R \to R[x_1, \dots, x_n]/(f_1, \dots, f_n)$ where ????

DEFINITION 5.1. *A map of k-schemes $f\colon X \to Y$ is* ***étale*** *if it is locally of finite type and if the following condition is satisfied. For every point $y \in Y$ and every $x \in X$ such that $f(x) = y$, there exist affine open sets $x \in U$ and $y \in V$ such that $f(V) \subseteq U$ and the map $f|_V \colon V \to U$ is isomorphic to spec of a standard étale map of rings.*

5.2. Rigid covers in algebraic geometry. The idea behind the definition of rigid covers in algebraic geometry is very similar to what we did in topology: a rigid étale cover is an étale cover $E \to X$ together with a choice, for every point in X, of a lifting into E. The only subtlety lies in our interpretation of the word "point".

DEFINITION 5.3. *Let X be a scheme. A* ***rigid étale cover*** *of X is an étale cover $E \to X$ together with, for every point $x \in X$, a map $\chi_x\colon \operatorname{Spec} \overline{k(x)} \to E$ making the following diagram commute:*

$$\begin{tikzcd} & & E \arrow[d] \\ \operatorname{Spec} \overline{k(x)} \arrow[r] \arrow[urr, "\chi_x"] & \operatorname{Spec} k(x) \arrow[r] & X. \end{tikzcd}$$

A map of rigid étale covers $E \to F$ is a map of étale covers such that for each $x \in X$ the diagram

$$\begin{tikzcd} E \arrow[rr] & & F \\ & \operatorname{Spec} \overline{k(x)}. \arrow[ul] \arrow[ur] & \end{tikzcd}$$

is commutative.

THEOREM 5.4. *Let X be a scheme. The category of rigid étale covers of X is cofiltered.*

6. Systems of approximations

In the past few sections we have dealt with various diagrams $I \to s\mathcal{S}et$, for different indexing categories I. We have been regarding such things as systems of combinatorial approximations to a homotopy type. So far, however, we have not formally addressed the issue of how to manipulate such systems. That is our goal in the present section.

6.1. Recovering topological invariants. Let X be a space. Suppose one has a pro-space $Z\colon I \to \mathcal{T}op$ together with a collection of compatible maps $Z_i \to X$. Assume that for each i in I and each $n \geq 0$, there is a j in I and a map $j \to i$ such that the map $Z_j \to X$ is an n-equivalence. Intuitively, this says that as one moves "outward" in the pro-space the spaces Z_i become better and better approximations to X. Under these conditions, what topological invariants of X can be recovered from the system Z?

The first thing to notice is that we can certainly recover the cohomology of X. Indeed, since an $(n+1)$-equivalence induces isomorphisms on $H^n(-)$, it follows readily that the canonical maps

$$H^n(X; A) \to \operatorname*{colim}_i H^n(Z_i; A)$$

are isomorphisms for any abelian group A. The situation is different for other cohomology theories, however. The group $K^0(X)$ can usually not be recovered from just knowing the n-type of X, no matter how large we take n to be. So the map

$$K^0(X) \to \operatorname*{colim}_i K^0(Z_i)$$

will generally not be an isomorphism. It is still possible to recover K-theoretic information about X from the pro-space Z, but it requires a more complicated technique. We describe this next.

Suppose W is a space which sits inside a homotopy fiber sequence

$$E_1 \to W \to E_2$$

where E_1 and E_2 are two Eilenberg-MacLane spaces. For any space Y there is an induced homotopy fiber sequence $F(Y, E_1) \to F(Y, W) \to F(Y, E_2)$ and therefore a long exact sequence of homotopy groups. Consider the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & \pi_k F(X, E_1) \arrow[r] \arrow[d] & \pi_k F(X, W) \arrow[r] \arrow[d] & \pi_k F(X, E_2) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & \operatorname*{colim}_i \pi_k F(Z_i, E_1) \arrow[r] & \operatorname*{colim}_i \pi_k F(Z_i, W) \arrow[r] & \operatorname*{colim}_i \pi_k F(Z_i, E_2) \arrow[r] & \cdots
\end{tikzcd}$$

Both the horizontal rows are exact, where for the bottom row we are using the fact that our indexing category is filtered. By what we have already remarked, the vertical maps where the codomain is either E_1 or E_2 are isomorphisms. So by the five-lemma, we find that

$$\pi_k F(X, W) \to \operatorname*{colim}_i \pi_k F(Z_i, W)$$

is an isomorphism as well. By the evident induction, we see that this works whenever W has a finite Postnikov tower—or said differently, whenever W has only finitely many nonzero homotopy groups.

The conclusion of the above paragraph can be improved. If W has a finite Postnikov tower, we saw that the homotopy groups $\pi_k F(X,W)$ can be recovered from the groups $\pi_k F(Z_i,W)$. Even more is true, though: the homotopy type of $F(X,W)$ can be recovered from the homotopy types of $F(Z_i,W)$: the map

$$F(X,W) \to \operatorname*{hocolim}_i F(Z_i,W)$$

is a weak equivalence.

Now let W be an arbitrary space (for example, $\mathbb{Z} \times BU$), and let

$$\cdots \to P_2W \to P_1W \to P_0W$$

be the Postnikov tower. By what has just been said, the maps

$$F(X,P_nW) \to \operatorname*{hocolim}_i F(Z_i,P_nW)$$

are weak equivalences for all n. But the groups in the domain and codomain both form towers, and taking the homotopy limit of both sides therefore gives a weak equivalence

$$\operatorname*{holim}_n F(X,P_nW) \xrightarrow{\sim} \operatorname*{holim}_n [\operatorname*{hocolim}_i F(Z_i,P_nW)].$$

Finally, we recall that $\operatorname{holim}_n F(X,P_nW) \simeq F(X,\operatorname{holim}_n P_nW) \simeq F(X,W)$. So we have the formula

$$F(X,W) \simeq \operatorname*{holim}_n [\operatorname*{hocolim}_i F(Z_i,P_nW)],$$

which holds for arbitrary spaces W.

So the conclusion is that when $p \geq 0$ we *can* recover the groups $K^{-p}(X)$ from the information in the pro-space Z, but it has to be done by a slightly complicated formula:

$$K^{-p}(X) = \pi_p F(X,\mathbb{Z}\times BU) \cong \pi_p\left(\operatorname*{holim}_n \left[\operatorname*{hocolim}_i F(Z_i,P_n(\mathbb{Z}\times BU))\right]\right).$$

By using function spaces in the category of spectra, we can do something similar for any connective cohomology theory E: for all $p \in \mathbb{Z}$,

$$E^p(X) \cong \pi_{-p}\left(\operatorname*{holim}_n \left[\operatorname*{hocolim}_i F(Z_i,P_nE)\right]\right).$$

The assumption that E is connective is needed to ensure that the Postnikov sections are finite extensions of Eilenberg-MacLane spectra.

6.2. Invariants of systems. We now adopt the following point of view. If E is a connective cohomology theory and $Z\colon I \to s\mathcal{S}et$ is a functor where I is cofiltered, then we *define* the E-cohomology of Z by the formula

$$E^p(Z) = \pi_{-p}\left(\operatorname*{holim}_n \left[\operatorname*{hocolim}_i F(Z_i,P_nE)\right]\right).$$

This is obviously inspired by the considerations of the previous section.

Likewise, we can define a mapping space $\operatorname{Map}(Z,E)$ by

$$\operatorname{Map}(Z,E) = \operatorname*{holim}_n \left[\operatorname*{hocolim}_i F(Z_i,P_nE)\right].$$

So $E^p(Z)$ is just the $-p$th cohomology group of $\operatorname{Map}(Z,E)$.

6.3. Comparing two systems. Let $X\colon I \to s\mathcal{S}et$ be a functor, where I is cofiltered. Let J be another cofiltered category, and let $\gamma\colon J \to I$ be a functor. We wish to compare X to the composite functor $X\gamma$.

6.4. The model category structure on pro-spaces.

7. Hypercovers and étale homotopy types

In this section we are finally able to define the étale realization of a scheme X. This requires that we introduce one last piece of machinery, however.

7.1. Hypercovers. Let $E \to X$ be an étale cover, and consider the associated Čech complex $\check{C}(E)$. Refining the cover to $E' \to X$ also gives us a refinement $\check{C}(E')_k \to \check{C}(E)_k$ of the k-fold pullbacks (for any k)—however, it does not give us an *arbitrary* refinement of $\check{C}(E)_k$. In other words, if $U \to \check{C}(E)_k$ is a cover then it is not clear that there is a refinement $E' \to E$ such that $\check{C}(E')_k \to \check{C}(E)_k$ factors through U. This is a slight obstacle in our overall plan, since our hope is for a system of approximations in which the schemes at each level in some sense become smaller and smaller.

The idea for fixing this problem leads at once to *hypercovers*. Essentially, these are simplicial schemes which are similar to Čech complexes but where in each level n one is allowed to further refine the n-fold pullbacks. To rigorously describe this we need a little machinery.

Recall that the cosimplicial indexing category Δ is the category whose objects are the sets $[n]$ and whose morphisms are the monotone increasing maps. Let Δ_+ be the **augmented simplicial category**, obtained by adding the empytset as an inital object. A functor $X\colon \Delta_+^{op} \to \mathcal{C}$ is called an augmented simplcial object in $\mathcal{C}$, and $X(\emptyset)$ is called the augmentation. It is useful to also use the notation $[-1]$ for the initial object of Δ_+.

For any simplicial set K, regard K as an augmented simplicial object by setting the augmentation to be a single point.

Let $\mathcal{C}$ be a category with limits. If $X \in \mathcal{C}$ and S is a set, then let X^S denote a product of copies of X indexed by the set S. For $U\colon \Delta_+^{op} \to \mathcal{C}$ be an augmented simplicial object. For any $K \in sSet$, define

$$\hom_+(K,U) = \mathrm{eq}\left[\prod_{n\geq -1} U_n^{K_n} \rightrightarrows \prod_{[k]\mapsto[m]} U_m^{K_k}\right].$$

It is easy to check that there are natural isomorphisms $\hom_+(\Delta_n, U) \cong U_n$. The object $\hom_+(\partial\Delta^n, U)$ is denoted M_nU and called the ***n*th (augmented) mathching object of *U***. The map

$$U_n = \hom_+(\Delta^n, U) \to \hom_+(\partial\Delta^n, U) = M_nU$$

is called the ***n*th matching map** of U.

EXERCISE 7.2. Let $E \to X$ be a map of schemes, and consider the simplicial scheme $\check{C}(E)$ which we regard as augmented by X. Then the maps $\check{C}(E)_n \to M_n\check{C}(E)$ are isomorphisms, for all n. Conversely, if $U \to X$ is an augmented simplicial scheme such that $U_n \to M_nU$ is an isomorphism for every n, then U is isomorphic to the Čech complex of $U_0 \to X$.

Definition 7.3. *Let X be a scheme. A **hypercover** of X is an augmented simplicial scheme $U_\bullet \to X$ such that for each n the augmented matching map $U_n \to M_n U$ is an étale cover.*

This, then, is the difference between hypercovers and Čech complexes. In the latter, each level n is precisely equal to the nth matching object (which intuitively consists of the n-fold intersections, or n-fold pullbacks, of the previous levels). In a hypercover one starts with this nth matching object but then is allowed to take a cover of it, and this may be done at each level.

Let $\mathrm{EtHyp}(X)$ denote the category of étale hypercovers of X. A map of hypercovers is simply a map of augmented simplicial schemes which is the identity on the augmentation. Just like the category of étale covers of X, the category $\mathrm{EtHyp}(X)$ is not cofiltered. To obtain a cofiltered category we need to use rigid covers.

Definition 7.4. *A **rigid hypercover** of a scheme X is a hypercover $U \to X$ together with the structure of rigid cover on every matching map $U_n \to M_n U$. A map of rigid hypercovers is a map of hypercovers which is compatible with the rigid structure in the evident way.*

Exercise 7.5.

(a) An ultra-rigid hypercover is a rigid hypercover $U \to X$ such that each matching map $U_n \to M_n U$ are ultra-rigid. If $U \to X$ is ultra-rigid and $V \to X$ is any rigid hypercover, check that there is at most one map of rigid hypercovers $U \to X$.

(b) Prove that the category $\mathrm{RgdEtHyp}(X)$ is cofiltered, and that the forgetful functor $\mathrm{RgdEtHyp}(X) \to \mathrm{EtHyp}(X)$ is final.

7.6. Étale homotopy types (finally). We can finally define the étale homotopy type of a scheme X. The phrase "étale homotopy type" is somewhat misleading, though, as it suggests an object in a homotopy category. As always, it is more convenient to work with a corresponding object in some underlying model category—we will call this object the "étale realization" of X.

Recall the following categories:

- The category $\mathrm{EtCov}(X)$ of étale covers of X.
- The category $\mathrm{RgdEtCov}(X)$ of rigid étale covers of X.
- The category $\mathrm{EtHyp}(X)$ of étale hypercovers of X.
- The category $\mathrm{RgdEtHyp}(X)$ of rigid étale hypercovers of X.

The second and fourth of these categories are cofiltered.

Definition 7.7. *The **étale realization** of a scheme X is the functor*

$$\mathrm{Et}_X\colon \mathrm{RgdEtHyp}(X) \to s\mathcal{S}et$$

which sends a rigid étale hypercover $U_* \to X$ to the simplicial set $\pi_0(U_*)$. We regard Et_X as an object in the category pro-s$\mathcal{S}$et.

Recall that pro-s$\mathcal{S}$et has a model category structure defined by Isaksen. The image of Et_X in Ho (pro-s$\mathcal{S}$et) is the **étale homotopy type** of X.

Remark 7.8. In early work on this subject, the correct model category structure on pro-s$\mathcal{S}$et was not available. Instead of using Ho (pro-s$\mathcal{S}$et), sources such as [**AM**] defined the étale homotopy type to be an object in $pro - \mathrm{Ho}\,(s\mathcal{S}et)$. This is the image of *our* étale homotopy type under a canonical functor forgetful functor

$\mathrm{Ho}\,(\text{pro-s}\mathcal{S}\text{et}) \to pro-\mathrm{Ho}\,(s\mathcal{S}et)$. In modern times it seems to be much more advantageous to work with $\mathrm{Ho}\,(\text{pro-s}\mathcal{S}\text{et})$, as here one has the underlying model structure on pro-s$\mathcal{S}$et available as a useful tool.

8. Étale cohomology and étale K-theory

CHAPTER 5

Sheaves and homotopy theory

CHAPTER 6

Topological interlude: Lefschetz pencils

The theory of Lefschetz pencils—and the closely related concept of Lefschetz fibrations—is based on a simple idea. Let $f\colon E \to B$ be a smooth map between real manifolds, and assume that the fibers are compact. As b varies inside of B, the fibers $f^{-1}(b)$ mostly have the same homotopy type: the homotopy type only changes when b is a critical value of f. Compactness of the fibers is important here; just consider the map $(S^1 \times I)\backslash\{(1, \frac{1}{2})\} \to I$ which projects onto the second coordinate! Now let us assume that f has only finitely many critical points, and that we understand the local behavior of f around each of these points: for instance, let us say that within a neighborhood of each critical point we can understand how the homotopy type of the fibers is changing. Then by local-to-global principles in homotopy theory (like the Mayer-Vietoris sequence) we can understand how the homotopy type of E is built from that of B.

In modern times the main application of these ideas is in Morse theory. There one studies maps $f\colon E \to \mathbb{R}$ having isolated and non-degenerate critical points—the so-called Morse functions. This means that in local coordinates around each critical point the function looks quadratic. Over the real numbers one precisely knows the different isomorphism classes of quadratic forms, and this reduces the local behavior to a discrete collection of possibilities. In Morse theory one learns that each critical point of f results in a change to the topology of E equivalent to attaching a cell, and the dimension of this cell is equal to the index of f near this critical point.

Lefschetz theory is very similar to Morse theory, and was developed around the same time. But here one looks at holomorphic maps of complex manifolds $f\colon E \to \mathbb{C}$ (or maps $E \to \mathbb{C}P^1$, which is almost the same), again with isolated and non-degenerate critical points. Over the complex numbers there is only one isomorphism class of nondegenerate quadratic form, namely the sum-of-squares mapping $(z_1, \ldots, z_n) \mapsto z_1^2 + \cdots + z_n^2$. This map has a single critical point at the origin, and it is not hard to analyze how the homotopy type of the fibers changes here—in fact it again amounts to adding a cell, this time always of dimension n. One can then use this information to understand the homotopy type of E.

The map f in the above paragraph is called a "Lefschetz fibration". What we have described so far amounts to a fairly simple idea. There is something subtle in the theory of Lefschetz fibrations that does not surface for Morse functions, however. A punctured disk in $\mathbb{C}$ has fundamental group $\mathbb{Z}$, and so there is a potential "twisting" in the fibers of f as they move around a critical point. This is called monodromy, and it influences the topology of E. The most difficult material in this chapter centers around such monodromy calculations. For Morse functions $f\colon E \to \mathbb{R}$ there is no analog to this, as one cannot move *around* the critical values.

Now suppose given an algebraic variety $X \hookrightarrow \mathbb{C}P^n$. Lefschetz's idea was to study hyperplane sections $X \cap H$ for various hyperplanes H in $\mathbb{C}P^n$. One readily sees that for most hyperplanes the intersections $X \cap H$ are homeomorphic to each other, but when H becomes tangent to X the homotopy type of $X \cap H$ changes. This information can be organized so that the intersections $X \cap H$ are the fibers of a map $E \to \mathbb{C}P^1$, with the base space $\mathbb{C}P^1$ parameterizing the hyperplanes being used. By carefully chosing these hyperplanes one can ensure that $E \to \mathbb{C}P^1$ has isolated and non-degenerate critical points, and then one is in the domain of ideas discussed above.

These methods allowed Lefschetz to develop a surprising amount of knowledge about the homology and cohomology groups of smooth, projective algebraic varieties. In Chapter 2 we met the Weak Lefschetz Theorem and the Hard Lefschetz Theorem, and in the present chapter we will describe how these tie in to the study of Lefschetz pencils. We will see that the proof of the Weak Lefschetz theorem is fairly easy and geometric. It seems that Lefschetz thought he had a geometric proof of the Hard Lefschetz Theorem, but in modern times no one has been able to understand this. The only known proofs of Hard Lefschetz are via Hodge theory or via Deligne's proof of the Riemann Hypothesis over finite fields! Obtaining a purely geometric proof of this theorem remains a tantalizing problem.

The best modern source for learning about Lefschetz's methods is a wonderful paper by Lamotke [**La**]. Our treatment follows this paper very closely. Lamotke's paper is itself much influenced by the work in [**SGA7b**].

1. Background

Before jumping into the full theory of Lefschetz pencils, it is useful to look carefully at some examples in the lowest dimension. This will serve to establish some intuition, and will also set the context of what was known before Lefschetz's work.

Let X be a projective algebraic curve over $\mathbb{C}$ (or equivalently, a compact Riemann surface). There is an old technique for understanding the homotopy type of X by examining a branched cover $p\colon X \to \mathbb{C}P^1$; this is a map that is *almost* a covering space, except that at certain "ramified points" the sheets of the cover come together. Said differently, the fibers of the map p mostly have the same cardinality, but there are certain points in $\mathbb{C}P^1$ where the cardinality drops (as the sheets meet). By examining the branching points in the cover, one can determine the Euler characteristic—and hence the genus—of X. In this section we review this classical technique and explain how it can be approached using pencils of hyperplanes.

1.1. A motivating example. Let $X \hookrightarrow \mathbb{C}P^2$ be the projective variety defined by the equation $x^3 + y^3 + z^3 = 0$. One readily checks that this is smooth, so topologically it is a compact 2-manifold. Which one is it? It will certainly be orientable, as X has a complex structure. So X is either S^2 or a genus g torus, and we can decide which by computing the Euler characteristic.

To do this, consider the map $p\colon X \to \mathbb{C}P^1$ defined by $p([x : y : z]) = [x : y]$. Note that this is well-defined, as X does not contain the point $[0 : 0 : 1]$. The reader will note that p is almost a covering space, but not quite: most fibers of p have exactly three points, but some have fewer. Without making any formal definitions, let us say that p is an example of a **branched cover** of degree 3. The set

$$B = \{q \in \mathbb{C}P^1 \mid p^{-1}(q) \text{ has fewer than 3 points}\}$$

is called the **branch locus** of p. A point $x \in X$ is an **unramified point** if p is a local homeomorphism near x, and otherwise x is called a ramified point. Finally, the **ramification set** R of the branched cover is the set of ramified points in X. Note that R is not necessarily equal to $p^{-1}(B)$, although in most cases we consider this will be true.

EXAMPLE 1.2. A main example to consider here is that of $P\colon \mathbb{C} \to \mathbb{C}$ given by $P(z) = z^n$. This is a branched cover of degree n, the branch locus is $\{0\}$, and the ramification set is also $\{0\}$.

In our example of $p\colon X \to \mathbb{C}P^1$, the branch locus is

$$B = \big\{[x : y] \mid x^3 + y^3 = 0\big\} = \big\{\, [1 : -1],\, [1 : -\zeta],\, [1 : -\bar{\zeta}]\big\}$$

where ζ is a primitive cube root of 1. The fiber over each of these points is a singleton, and so the ramification set also consists of three points.

Let U_1, U_2, and U_3 be small Euclidean neighborhoods around each of the ramification points r_1, r_2, and r_3. Let U be their (disjoint) union. Then

$$\begin{aligned}\chi(X) = \chi(X-R) + \chi(X, X-R) &= \chi(X-R) + \chi(U, U-R) \qquad \text{(excision)}\\ &= \chi(X-R) + \sum_i \chi(U_i, U_i - \{r_i\})\\ &= \chi(X-R) + \sum_i \big[\chi(U_i) - \chi(U_i - \{r_i\})\big]\\ &= \chi(X-R) + \sum_i \big[1 - \chi(U_i - \{r_i\})\big].\end{aligned}$$

But the restriction of p to $X - R \to \mathbb{C}P^1 - B$ is a covering space of degree 3, so we have that

$$\chi(X-R) = 3 \cdot \chi(\mathbb{C}P^1 - B) = 3 \cdot \chi(S^2 - \{3 \text{ points}\}) = 3 \cdot -1 = -3.$$

Write $b_i = p(r_i)$. Then the restriction of p to $U_i - \{r_i\} \to p(U_i) - b_i$ is likewise a covering space, so we have

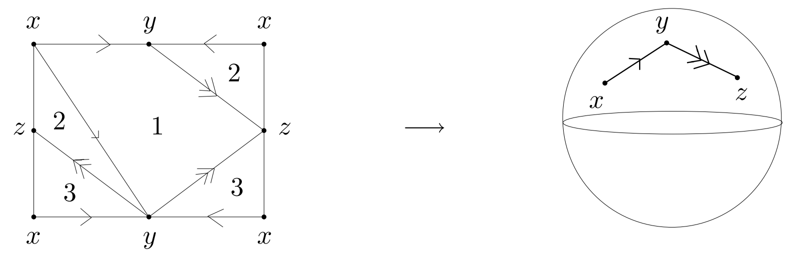
$$\chi(U_i - \{r_i\}) = 3 \cdot \chi(p(U_i) - b_i) = 3 \cdot 0 = 0.$$

Substituting into the formula for $\chi(X)$, we find that

$$\chi(X) = -3 + 3 = 0.$$

Hence, topologically X is a torus.

The reader might wonder exactly how a torus can map to S^2 as a degree 3 branched cover. At first this is hard to picture! The following diagram demonstrates such a mapping, where the torus is modelled by a square with opposite edges identified as usual (although these identification are *not* indicated in the picture):



To understand what is being shown here, first imagine making two adjoining incisions in S^2, with cut points x, y, and z as shown. If you unfold S^2 along these incisions you get a quadrilateral, and we imagine marking the edges to indicate how they should be glued back together. We now draw lines on the torus to break it up into three regions, marked 1, 2, and 3 in the diagram, and we label the edges of these regions to match the labelling on our quadrilateral. Choose orientations on the torus and the sphere, and map each region of the torus into our sphere in such a way that the arrows match up and the orientations match up (this is easiest to see in the case of region 1, whereas for the other regions one has to mentally remove their pieces from the picture and rearrange them to look like quadrilaterals). This gives us a mapping from $T \to S^2$ which is a 3-fold cover over all points except for x, y, and z, and the preimage of each of these points is a singleton (shown in the diagram of the torus as labelled by the same letter).

1.3. The homotopy type of curves in $\mathbb{C}P^2$. Now, with a little modification one can apply the above method to determine the homotopy type of every smooth hypersurface in $\mathbb{C}P^2$. First let us generalize the Euler characteristic argument for branched covers.

THEOREM 1.4 (Riemann–Hurwitz Formula). *Let $p\colon X \to Y$ be a d-fold branched cover, where Y is an n-manifold and the branch locus B is finite. Then*

$$\chi(X) = d[\chi(Y) - \#B] + \#p^{-1}(B).$$

PROOF. For each point $x \in p^{-1}(B)$ we can associate a "local degree" e_x by looking at small neighborhoods V around x and taking the degree of the covering space $V - \{x\} \to p(V) - p(x)$. In fact this definition works for any $x \in X$. It is clear that

$$\sum_{x \in p^{-1}(y)} e_x = d$$

for each $y \in Y$.

Now let $x_1, \dots, x_t$ be the points in $p^{-1}(B)$, and choose small disjoint Euclidean neighborhoods U_i around the x_i's such that $p|_{U_i - \{x_i\}}$ is a covering space. Let e_i be the local degree of the cover near x_i. Arguing as before, we find that

$$\begin{aligned}
\chi(X) &= \chi(X - p^{-1}(B)) + \chi(X, X - p^{-1}(B)) \\
&= d \cdot \chi(Y - B) + \sum_i \chi(U_i, U_i - \{x_i\}) \\
&= d \cdot [\chi(Y) + (-1)^{n-1}\#B] + \sum_i [\chi(U_i) - \chi(U_i - \{x_i\})] \\
&= d \cdot [\chi(Y) + (-1)^{n-1}\#B] + \#p^{-1}(B) - \sum_i e_i \cdot \chi(p(U_i) - p(x_i)) \\
&= d \cdot [\chi(Y) + (-1)^{n-1}\#B] + \#p^{-1}(B) - \sum_i e_i \cdot \chi(S^{n-1}) \\
&= d \cdot [\chi(Y) + (-1)^{n-1}\#B] + \#p^{-1}(B) - \sum_i e_i [1 + (-1)^{n-1}] \\
&= d \cdot [\chi(Y) + (-1)^{n-1}\#B] + \#p^{-1}(B) - d \cdot \#B \cdot [1 + (-1)^{n-1}] \\
&= d[\chi(Y) - \#B] + \#p^{-1}(B).
\end{aligned}$$

□

Suppose $X \hookrightarrow \mathbb{C}P^2$ is a smooth subvariety defined by the homogeneous degree d equation $f(x, y, z) = 0$. We can assume that f has a term of the form az^d where $a \in \mathbb{C} - \{0\}$, since if not we could perform a linear change of variables in x, y, and z to assure that this is the case. The presence of the az^d term guarantees that X does not contain the point $[0 : 0 : 1]$, and so we can once again consider the map $p\colon X \to \mathbb{C}P^1$ given by $[x : y : z] \mapsto [x : y]$. This is a degree d branched cover.

To use the Riemann-Hurwitz formula to compute $\chi(X)$, we need an understanding of the branch locus. That is, for what values of $[x : y]$ will there be fewer than d roots of the equation $f(x, y, z) = 0$? Said differently, for fixed values of x and y the equation $f(x, y, z) = 0$ becomes a polynomial equation of degree d in the single variable z. The presence of repeated roots for this polynomial is governed by an algebraic expression in the coefficients called the *discriminant*. We pause here to review this piece of algebra.

1.5. Review of the discriminant. A quadratic polynomial $p(x) = ax^2 + bx + c$ has a repeated root if and only if $b^2 - 4ac = 0$. In other words, there is a polynomial expression in the coefficents—called the discriminant—whose vanishing is equivalent to $p(x)$ having a repeated root. It turns out that this is true for polynomials of arbitrary degree, although the actual form of the discriminant becomes quite complicated.

Let $p(x) = (x - r_1)(x - r_2) \cdots (x - r_d)$ where both x and the r_i's are indeterminants. That is, think of $p(x)$ as an element of $k[x, r_1, \dots, r_d]$. Then

$$p(x) = x^d + \sigma_1(r)x^{d-1} + \cdots + \sigma_{d-1}(r)x + \sigma_d(r)$$

where the σ_i's are the elementary symmetric functions in d variables. Consider the expression

$$D = \prod_{i<j}(r_i - r_j)^2 \in k[r_1, \dots, r_d],$$

and note that it is homogeneous in the r_i's of degree $2 \cdot \binom{d}{2} = d(d-1)$. This expression is invariant under permutation of the r_i's, hence there exists a unique polynomial $\Delta_d(w_1, \dots, w_d)$ such that

$$D = \Delta_d(\sigma_1(r), \dots, \sigma_d(r)).$$

The polynomial $\Delta_d(w_1, \dots, w_d)$ is called the **universal discriminant polynomial** for degree d. If w_i is regarded as having degree i, then Δ_d is homogeneous of degree $d(d-1)$.

Suppose given a polynomial $p(x) = a(x^d + b_1 x^{d-1} + \cdots + b_{d-1}x + b_d)$, where $a, b_i \in k$. We define it's discriminant to be $a^{2d-2}\Delta_d(b_1, b_2, \dots, b_d)$. It is evident that $p(x)$ has a repeated root if and only if the discriminant is zero.

EXAMPLE 1.6. When $d = 2$ then we have $D = [r_1 - r_2]^2 = (r_1 + r_2)^2 - 4r_1 r_2 = \sigma_1^2 - 4\sigma_2$. So $\Delta_2(w_1, w_2) = w_1^2 - 4w_2$. Given a polynomial $p(x) = ax^2 + bx + c$, we write it as $p(x) = a[x^2 + \frac{b}{a}x + \frac{c}{a}]$ and then it's discriminant is

$$a^2\Delta\Big(\frac{b}{a}, \frac{c}{a}\Big) = a^2\Big[\frac{b^2}{a^2} - 4\frac{c}{a}\Big] = b^2 - 4ac.$$

EXAMPLE 1.7. For a cubic polynomial $p(x) = ax^3 + bx^2 + cx + d$, the discriminant is $b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$. For quartic polynomials the discriminant is too unpleasant to write down!

1.8. Back to the main argument. Recall our basic setup. We are considering a smooth subvariety $X \hookrightarrow \mathbb{C}P^2$ defined by a homogeneous degree d equation $f(x, y, z) = 0$. We can assume that f has a term of the form az^d where $a \in \mathbb{C} - \{0\}$, and so we have the map $p\colon X \to \mathbb{C}P^1$ given by $[x : y : z] \mapsto [x : y]$. This is a degree d branched cover.

Now write $f(x, y, z) = az^d + f_1(x, y)z^{d-1} + f_2(x, y)z^{d-2} + \cdots + f_d(x, y)$ where each $f_i(x, y)$ is homogeneous of degree 1. Then the branch locus of p is the vanishing set of the polynomial

$$\Delta_d(f_1(x, y), \dots, f_d(x, y)) \tag{1.9}$$

which is a degree $d(d-1)$ homogeneous polynomial in x and y. Generically (i.e., for all choices of f outside a set of measure zero), the vanishing set will be exactly $d(d-1)$ points in $\mathbb{C}P^1$.

To understand the set $p^{-1}(B)$, we again think generically. Given a point $[x : y]$ in B, the fiber of p over $[x : y]$ will be in bijective correspondence with the roots z of the equation

$$az^d + f_1(x, y)z^{d-1} + f_2(x, y)z^{d-2} + \cdots + f_d(x, y) = 0.$$

There are at most d of these roots, and the fact that $[x : y]$ lies in B is equivalent to saying that there are *fewer* then d roots. For generic choices of f we can guarantee that there will be exactly $d - 1$ roots for each $[x : y]$, and so $p^{-1}(B)$ will consist of $\#B \cdot (d - 1) = d(d - 1)^2$ elements. The Riemann–Hurwitz formula then gives us that

$$\begin{aligned}\chi(X) = d \cdot [\chi(\mathbb{C}P^1) - d(d-1)] + d(d-1)^2 &= 2d - d^2(d-1) + d(d-1)^2 \\ &= 2d - d(d-1) \\ &= 3d - d^2.\end{aligned}$$

If we want the genus of X, it is

$$g = \frac{2 - \chi(X)}{2} = \frac{d^2 - 3d + 2}{2} = \binom{d-1}{2}.$$

To analyze the non-generic case we need more sophisticated algebra, and we only sketch this. If (1.9) has fewer than $d(d-1)$ roots, this coincides exactly with the case where the fibers over some points in the branch locus have fewer than $d-1$ points. A double root of (1.9) means the corresponding fiber has $d-2$ points, a triple root means $d-3$ points, and so on. This says that if the branch locus has $d(d-1)-r$ elements, then the number of elements in $p^{-1}(B)$ is

$$[d(d-1)-r]\cdot(d-1)-r.$$

This number comes about by thinking of each branch point as generically having $d-1$ points in its fiber, and then reducing the number by 1 for each repeated root of (1.9). By the Riemann–Hurwitz formula we find that

$$\chi(X) = d[2-[d(d-1)-r]] + [d(d-1)-r](d-1)-r = 3d-d^2,$$

which is the same result as before. This coincidence may seem surprising, as it has come out almost accidentally from our work. Later we will see how we could have guessed this in the first place. In any case, we have proven the following:

THEOREM 1.10. *A smooth hypersurface of degree d in $\mathbb{C}P^2$ is homeomorphic to a torus of genus $\binom{d-1}{2}$.*

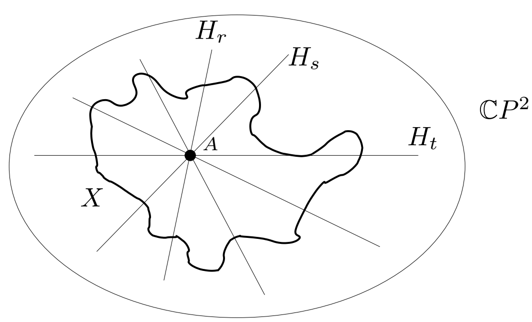
1.11. Another perspective. For each $x, y \in \mathbb{C}$ except $x = y = 0$, let $f_{x,y}\colon \mathbb{C}P^1 \hookrightarrow \mathbb{C}P^2$ be the linear embedding

$$[u:v] \mapsto [xu:yu:v]$$

and let $H_{x,y}$ be its image. Note that $H_{x,y}$ only depends on the point $[x:y]$ in $\mathbb{C}P^1$, and it can also be described as the hyperplane defined by the linear equation $yz_1 - xz_2 = 0$ (where $\mathbb{C}P^2$ has homogeneous coordinates z_1, z_2, z_3). It is easy to check that:

- Each $H_{x,y}$ contains the point $[0:0:1]$;
- For $[x:y] \neq [u:v]$ one has $H_{x,y} \cap H_{u,v} = \{[0:0:1]\}$.
- Except for $[0:0:1]$, each point of $\mathbb{C}P^2$ lies on a unique H_t for some $t = [x:y] \in \mathbb{C}P^1$.

The collection of all the $H_{x,y}$'s is called a **pencil of hyperplanes**, and the point $A = [0:0:1]$ is called its **axis**. One should imagine the following schematic picture:



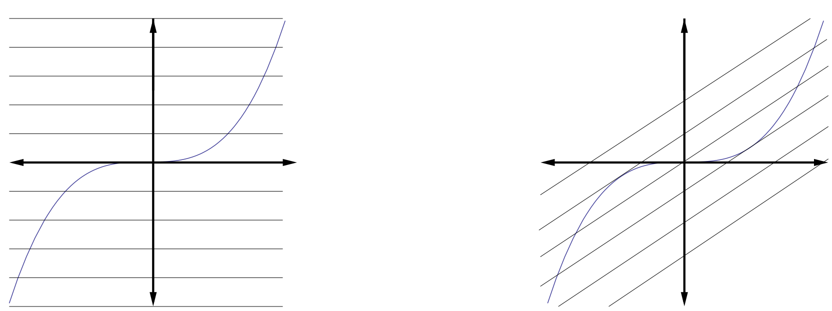
Provided that X does not contain $[0:0:1]$, we can define a map $p\colon X \to \mathbb{C}P^1$ by sending each point x to the unique t such that x lies in H_t. In terms of a formula, this is exactly $[x:y:z] \mapsto [x:y]$. The fibers of this map are just the intersections $X \cap H_t$, for different values of $t \in \mathbb{C}P^1$. For most values of t the set $X \cap H_t$ consists of exactly d points—these are the t where X intersects H_t transversally. For some values of t the projective line H_t is tangent to X, and when this happens there will be fewer than d points of intersection, depending on the degree of tangency.

Of course there is nothing special about the point $[0 : 0 : 1]$ here. If X did happen to contain this point, we could choose another point (not on X) to be the axis of our pencil. This would result in a different formula for the map p, but the geometry is exactly the same. In fact we could even perform a linear change of coordinates in $\mathbb{C}P^2$ so that our axis became $[0 : 0 : 1]$ again. Recall that this is precisely what we did back in Section 1.3, when we arranged for the polynomial f to have a z^d term.

The following fact is not really necessary for the arguments we have done in this section, but we state it here because it will be important later and because it is easy to understand at this simple stage. We have quite a bit of freedom in chosing our pencil of hyperplanes, and by moving it around we can make the "degenerate" intersections $X \cap H_t$ as nice as possible. We cannot remove tangencies altogether, but it turns out that we can eliminate degree 3 tangencies and higher by moving the pencil slightly. That is to say,

Fact: One can always find a pencil of hyperplanes such that each H_t intersects X at points with multiplicity at most 2—in other words, at points where either the intersection is transverse or where there is a simple tangency.

Pencils with the above property are called "Lefschetz pencils," as we will learn in subsequent sections. The above fact is certainly not obvious, but hopefully it seems believable. The following picture suggests the basic idea (despite depicting the situation over $\mathbb{R}$, and in the affine plane rather than projective space). On the left, X is the graph of the curve $y = x^3$ and we are looking at the pencil of horizontal hyperplanes, which are just lines in this case. The line $y = 0$ intersects X in a triple tangency. But by moving the pencil slightly—in this case, by rotating the lines of the pencil—one obtains a new pencil where the tangencies are now ordinary double points. In fact *almost all* pencils are Lefschetz pencils; tangencies of order higher than 2 only occur on a set of measure zero.



Summary: In this section we have described a method for understanding the homotopy type of smooth algebraic curves in $\mathbb{C}P^2$ which involves mapping them to $\mathbb{C}P^1$ and counting the points in the various fibers of this map. One can think of the map to $\mathbb{C}P^1$ as arising from slicing the curve by the hyperplanes in a chosen pencil for $\mathbb{C}P^2$. This technique is probably as old as topology itself, going back at least to Poincaré.

Lefschetz had the idea of extending this technique to higher dimensional varieties. If a smooth variety maps to $\mathbb{C}P^1$, the fibers are algebraic varieties of one dimension less. Most of the fibers are homeomorphic to each other, but there are some degenerate fibers where the topology changes. Lefschetz's program was to use these ideas to inductively study the topology of algebraic varieties. We will develop this program over the course of the next few sections.

2. The topology of the sum-of-squares mapping

Let S^n be the set of points $\mathbf{x} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$ such that $x_0^2 + \cdots + x_n^2 = 1$. This is the usual topological n-sphere. But now consider the space

$$S^n_{\mathbb{C}} = \{(z_0, \ldots, z_n) \in \mathbb{C}^{n+1} \mid z_0^2 + \cdots + z_n^2 = 1\}.$$

Note that $S^n_{\mathbb{C}}$ contains S^n as its real-valued points, but also that $S^n_{\mathbb{C}}$ is not compact: in fact, for any complex numbers $z_0, \ldots, z_{n-1}$ there exists at least one (and usually two) points of $S^n_{\mathbb{C}}$ having these numbers as its first n components. The first result we need is:

Proposition 2.1. *The inclusion $S^n \hookrightarrow S^n_{\mathbb{C}}$ admits a deformation retraction.*

Proof. If we write $z_j = x_j + iy_j$, then the equation $\sum z_j^2 = 1$ is equivalent to the two equations $\sum x_j^2 - \sum y_j^2 = 1$ and $\sum x_j y_j = 0$. That is, a point of $S^n_{\mathbb{C}}$ may be thought of as a pair of vectors $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{\mathbf{n}} \times \mathbb{R}^{\mathbf{n}}$ such that $|\mathbf{x}|^2 = |\mathbf{y}|^2 + 1$ and $\mathbf{x} \cdot \mathbf{y} = 0$. Under this model, $S^n \hookrightarrow S^n_{\mathbb{C}}$ is simply the subpace of pairs where $\mathbf{y} = 0$.

The retraction $S^n_{\mathbb{C}} \to S^n$ is given by $(\mathbf{x}, \mathbf{y}) \mapsto \frac{\mathbf{x}}{||\mathbf{x}||}$ (note that $\mathbf{x}$ is nonzero, as its norm is at least 1). The deformation is given by linearly shrinking $\mathbf{y}$ to zero while at the same time scaling $\mathbf{x}$ appropriately—we leave it to the reader to write down appropriate formulas. $\square$

Corollary 2.2. *For any nonzero complex number w, the affine variety $\{(z_0, \ldots, z_n) \in \mathbb{C}^n \mid \sum z_j^2 = w\}$ is homotopy equivalent to S^n.*

Proof. Choose a λ such that $\lambda^2 = w$. Then $(z_0, \ldots, z_n) \mapsto (z_0/\lambda, \ldots, z_n/\lambda)$ gives a homeomorphism betwen our variety and $S^n_{\mathbb{C}}$. $\square$

Next consider the map $f\colon \mathbb{C}^n \to \mathbb{C}$ given by $f(z_1, \ldots, z_n) = z_1^2 + \cdots + z_n^2$. For each $w \in \mathbb{C}$ let F_w denote the fiber $f^{-1}(w)$. By Corollary 2.2 we have that when w is nonzero the fiber F_w has the homotopy type of S^{n-1}. In fact it is easy to show that the restriction $f^{-1}(\mathbb{C} - 0) \to \mathbb{C} - 0$ is a fiber bundle. The fiber F_0 certainly does *not* have the homotopy type of S^{n-1}, though—in fact this fiber is readily seen to be contractible using the homotopy $(\mathbf{z}, t) \mapsto t\mathbf{z}$.

The map f is the archetype for what will be called a *Lefschetz fibration*, and we can already see the main point about them. As the generic fiber degenerates into the special fiber, the $(n-2)$-homotopy type is not changing. That is to say, the homotopy groups of the generic and special fibers agree in dimensions less than $n - 1$. Moreover, in dimension $n - 1$ we can also say exactly what is happening: a certain homotopy element of the generic fiber vanishes inside the special fiber. Readers who know some Morse Theory should see the parallels here: the special fiber is obtained from the nearby fibers by attaching an n-cell.

2.3. Vanishing cycles. For $w \in \mathbb{C} - 0$ the fiber F_w has the homotopy type of an $(n-1)$-sphere. A generator of $H_{n-1}(F_w)$ is called a **vanishing cycle** for f. The terminology comes from the fact that as F_w approaches the singular fiber (that is, as w approaches 0) the cycle shrinks and then ultimately vanishes. We can see this very precisely as follows. For $w \in \mathbb{C}$, let $\sqrt{w}$ denote a chosen square root of w. Let $j_w \colon S^{n-1} \to F_w$ be the map

$$(r_1, \dots, r_n) \mapsto (r_1\sqrt{w}, \dots, r_n\sqrt{w}).$$

We have seen that this is a homotopy equivalence, so applying $(j_w)_*$ to a chosen generator for $H_{n-1}(S^{n-1})$ yields a vanishing cycle. If we let w approach 0 along some path, choosing $\sqrt{w}$ continuously as we go, we see that the image of j_w shrinks in radius until we get to $w = 0$ where the radius vanishes and the image is just a point.

Let $E = \mathbb{C}^n$ and $B = \mathbb{C}$, so that our sum-of-squares map is $f \colon E \to B$. Another way to express the "vanishing" aspect of the vanishing cycles is to say that if $i \colon F_w \hookrightarrow E$ is the inclusion then the map $i_* \colon H_*(F_w) \to H_*(E)$ takes the vanishing cycles to zero. Now, in the present example this is a silly statement because E is contractible—but later we will see contexts where this phrasing takes on more significance. We mention it here only to tie it in with the previous paragraph. If we let w approach zero by the straight-line path then we can write down a map $D^n \to E$ via the formula

$$t(r_1, \dots, r_n) \mapsto (r_1\sqrt{tw}, \dots, r_n\sqrt{tw})$$

for $(r_1, \dots, r_n) \in S^{n-1}$. This is called the **thimble** associated to the vanishing cycle j_w. If one draws a picture then the image of D^n looks like a cone, or a (very pointy) thimble. The thimble gives an explicit null-homotopy for the composition $S^{n-1} \to F_w \hookrightarrow E$, or an explicit way of seeing that the vanishing cycle becomes a boundary inside of E.

The vanishing cycle and thimble together give maps

$$\begin{tikzcd} S^{n-1} \arrow[r, "\delta"] \arrow[d, tail] & F_w \arrow[d, tail] \\ D^n \arrow[r, "\Delta"] & E \end{tikzcd}$$

and therefore a map $H_n(D^n, S^{n-1}) \to H_n(E, F_w)$. One readily checks that this is an isomorphism. The domain is $\mathbb{Z}$, and a choice of orientation for D^n determines a generator.

2.4. The bounded sum-of-square mapping. Fix a real number $r > 0$. Let $S^n(r)$ be the usual sphere of radius r in $\mathbb{R}^{n+1}$, i.e., the set of points $(x_0, \dots, x_n) \in \mathbb{R}^{n+1}$ such that $\sum x_j^2 = r^2$.

Now fix $\rho > 0$ and $\epsilon > 0$ such that $\rho < \epsilon^2$. Let

$$E = \Big\{ (z_1, \dots, z_n) \in \mathbb{C}^n \;\Big|\; \sum_j |z_j|^2 \le \epsilon^2 \text{ and } |z_1^2 + \cdots + z_n^2| \le \rho \Big\}$$

and

$$B = \big\{ w \in \mathbb{C} \;\big|\; |w| \le \rho \big\}.$$

The sum-of-squares map $f \colon \mathbb{C}^n \to \mathbb{C}$ restricts to a map $E \to B$, and we will refer to this restriction as a **bounded sum-of-squares mapping**. We'll denote this restriction as g. Note that the condition $\rho < \epsilon^2$ simply guarantees that g is surjective. The map g behaves in almost exactly the same way as the original map f. Namely,

(1) For every nonzero $w \in B$, the fiber $g^{-1}(w)$ is homotopy equivalent to an $(n-1)$-sphere. Indeed, when $w \in \mathbb{R}$ the inclusion $S^{n-1}(\sqrt{|w|}) \hookrightarrow g^{-1}(w)$ is a deformation retraction, and all the other fibers (except $w = 0$) are homeomorphic to this one.
(2) The fiber $g^{-1}(0)$ is contractible.
(3) The restriction $E - g^{-1}(0) \to B - 0$ is a fiber bundle.

These facts are easy exercises that we will leave to the reader.

One thing that is *different* about the present setting is that the fibers are now manifolds with boundary. As the fibers over all nonzero points are homemorphic, let us for convenience just consider the fiber over ρ:

$$F_\rho = \big\{(z_1, \dots, z_n) \bigm| \sum |z_j|^2 \le \epsilon^2 \text{ and } \sum z_j^2 = \rho\big\}$$
$$\partial F_\rho = \big\{(z_1, \dots, z_n) \bigm| \sum |z_j|^2 = \epsilon^2 \text{ and } \sum z_j^2 = \rho\big\}.$$

Writing each z_j as $x_j + iy_j$, we have

$$\partial F_\rho \cong \Big\{(x, y) \bigm| x, y \in \mathbb{R}^n,\ |x|^2 + |y|^2 = \epsilon^2,\ |x|^2 - |y|^2 = \rho,\ x \cdot y = 0\Big\}$$
$$\cong \Big\{(x, y) \bigm| x, y \in \mathbb{R}^n,\ |x|^2 = \frac{\epsilon^2 + \rho}{2},\ |y|^2 = \frac{\epsilon^2 - \rho}{2},\ x \cdot y = 0\Big\}.$$

From the last description one sees immediately that ∂F_ρ is homeomorphic to the Stiefel manifold $V_2(\mathbb{R}^n)$ of 2-frames in $\mathbb{R}^n$: to get the homeomorphism one only has to normalize each of x and y.

In fact the same homeomorphisms show that the pair $(F_\rho, \partial F_\rho)$ is homeomorphic to the pair $(D(TS^{n-1}), S(TS^{n-1}))$ consisting of the disk- and sphere-bundles of the tangent bundle to S^{n-1}. In terms of explicit formulas one could write

$$D(TS^{n-1}) = \big\{(u, v) \bigm| u, v \in \mathbb{R}^n,\ |u| = 1,\ |v| \le 1,\ u \cdot v = 0\big\}$$
$$S(TS^{n-1}) = \big\{(u, v) \bigm| u, v \in \mathbb{R}^n,\ |u| = 1,\ |v| = 1,\ u \cdot v = 0\big\}.$$

We will need to know some basic facts about the homology groups of ∂F_ρ and their relationship to the homology groups of F_ρ. These things turn out to depend on the parity of n. The necessary facts are summarized below:

(1) Projection onto the first vector $p\colon V_2(\mathbb{R}^n) \to S^{n-1}$ is a fiber bundle with fiber S^{n-2}. The Serre spectral sequence gives $H_0(V_2(\mathbb{R}^n)) = H_{2n-3}(V_2(\mathbb{R}^n)) = \mathbb{Z}$ and a single differential $\mathbb{Z} \to \mathbb{Z}$ whose kernel and cokernel are H_{n-1} and H_{n-2}, respectively. All other homology groups are zero.
(2) When $n \ge 4$ there is a CW-complex structure for $V_2(\mathbb{R}^n)$ in which the n-skeleton is homeomorphic to $\mathbb{R}P^{n-1}/\mathbb{R}P^{n-3}$ (see [**MT**, Corollary 1 of Chapter 5]). Consequently, the cellular chain complex for $V_2(\mathbb{R}^n)$ has

$$0 \longrightarrow \mathbb{Z} \xrightarrow{1+(-1)^{n-1}} \mathbb{Z} \longrightarrow 0$$

in dimensions n through $n - 3$.

(3) When $n \geq 3$ the groups $H_*(V_2(\mathbb{R}^n))$ in dimensions $n-1$ and $n-2$ are equal to

	n even	n odd
$n-1$	$\mathbb{Z}$	0
$n-2$	$\mathbb{Z}$	$\mathbb{Z}/2$

The case $n = 3$ is included in the table, but does not follow from (2)—it must be dealt with separately. Here $V_2(\mathbb{R}^3) \cong \mathbb{R}P^3$, the homeomorphism coming about in two steps. First, $V_2(\mathbb{R}^3) \cong SO(3)$ using the cross-product: the 2-frame (v, w) maps to the matrix with columns v, w, and $v \times w$. Next, a classical construction associates to a vector $v \in D^3$ the rotation of $\mathbb{R}^3$ about the line $\mathbb{R}.v$, through an angle of $\pi \cdot |v|$ radians, directed counterclockwise from the point of view of a person standing at the tip of v and looking towards the origin. This gives a map $D^3 \to SO(3)$ (well-defined even when $v = 0$) that identifies antipodal points on the boundary and gives a homeomorphism $\mathbb{R}P^3 \cong SO(3)$.

(4) We will also have need of the case $n = 2$. Here $V_2(\mathbb{R}^2) \cong O(2) \cong S^1 \amalg S^1$. So $H_0 = H_1 = \mathbb{Z} \oplus \mathbb{Z}$ in this case. Note that this may be regarded as continuing the pattern given in (3) for n even: the homology of $V_2(\mathbb{R}^n)$ has the 'standard' copies of $\mathbb{Z}$ in dimensions 0 and $2n-3$ and an 'extra' copy of $\mathbb{Z}$ in dimensions $n-1$ and $n-2$. When $n = 2$ it just happens that these extra copies lie in the same dimensions as the standard copies.

(5) Next we need to analyze the long exact homology sequence for the pair $(F_\rho, \partial F_\rho)$. Lefschetz Duality says $H_i(F_\rho, \partial F_\rho) \cong H^{2n-2-i}(F_\rho)$, and since $F_\rho \simeq S^{n-1}$ this group is nonzero only when $i = n-1$. The long exact homology sequence is therefore mostly trivial except for the following piece:

$$0 \to H_{n-1}(\partial F_\rho) \to H_{n-1}(F_\rho) \to H_{n-1}(F_\rho, \partial F_\rho) \to H_{n-2}(\partial F_\rho) \to 0.$$

The middle two groups are both $\mathbb{Z}$, and we have seen that the identity of the other groups depends on the parity of n. When n is even the sequence is

$$0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z} \to 0$$

which shows that the middle map is zero and the other two are isomorphisms. When n is odd the sequence is

$$0 \to 0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2 \to 0$$

which shows that the middle map is multiplication by 2.

(6) There are two particular pieces of geometric information we need to extract from the above sequences. First, when n is even the generator of $H_{n-1}(F_\rho)$ may be deformed to the boundary. Second, when n is odd the generator of $H_{n-1}(F_\rho, \partial F_\rho; \mathbb{Q})$ (note the rational coefficients!) may be deformed so that it lies entirely in F_ρ.

2.5. Monodromy and variation. We will need to also do a monodromy calculation, so let us review how this works. Suppose $p\colon E \to B$ is a fiber bundle, and assume for convenience that the fiber F is a CW-complex. Then for every $b \in B$ one obtains a map of groups from $\pi_1(B, b)$ into the group $\mathrm{SHE}(F_b) \subseteq [F_b, F_b]$ of self-homotopy equivalences of F_b. This is called the **monodromy action** of $\pi_1(B, b)$ on F_b. As an immediate corollary one obtains an induced action of $\pi_1(B, b)$ on topological invariants of F_b such as $H^*(F_b)$ and $H_*(F_b)$.

To construct the monodromy action, let $\gamma\colon I \to B$ be a loop at b (so $\gamma(0) = \gamma(1) = b$). Consider the digram

$$\begin{tikzcd} F \times \{0\} \arrow[r] \arrow[d, tail] & E \arrow[d, "p"] \\ F \times I \arrow[r] & B \end{tikzcd}$$

where the lower horizontal map is the composite $F \times I \to I \xrightarrow{\gamma} B$ and the upper horizontal map is the inclusion of the fiber over b. Since $E \to B$ is a fiber bundle, the square has a lifting $\lambda\colon F \times I \to E$. Restricting λ to $F \times \{1\}$ gives a map $\phi(\gamma)\colon F_b \to F_b$. One proves that

(1) A different choice of λ will not change the homotopy class of $\phi(\gamma)$;
(2) If γ' and γ represent the same element of $\pi_1(B, b)$, then the maps $\phi(\gamma)$ and $\phi(\gamma')$ are homotopic; and
(3) $\phi(\gamma \cdot \beta)$ is homotopic to $\phi(\gamma) \circ \phi(\beta)$.

Facts (1) and (2) are simple consequences of the homotopy lifting property for fiber bundles, and fact (3) follows immediately from the construction. Together, these facts tell us that we have a map of groups $\phi\colon \pi_1(B, b) \to \mathrm{SHE}(F_b)$.

Now we do a calculation. Let $f\colon \mathbb{C}^n \to \mathbb{C}$ be the sum-of-squares map, and let $Q = f^{-1}(0)$. Then the restriction $\mathbb{C}^n - Q \to \mathbb{C} - 0$ is a fiber bundle with fiber $S^{n-1}_{\mathbb{C}}$. Let $1 \in \mathbb{C} - 0$ be the basepoint. The fundamental group of the base is just $\mathbb{Z}$, and the cohomology of the fibers is interesting only in dimension $n - 1$. Our monodromy action on cohomology is a homomorphism $\mathbb{Z} \to \mathrm{Aut}(H^{n-1}(S^n_{\mathbb{C}}))$, which will be determined by the image of a generator.

Proposition 2.6. *Let γ be the generator $t \mapsto e^{2\pi i t}$ of $\pi_1(\mathbb{C} - 0, 1)$. Then the monodromy action of γ on $H^{n-1}(S^{n-1})$ is multiplication by $(-1)^n$. The same is true for the action on $H_{n-1}(S^{n-1})$.*

Proof. Certainly the action is multiplication by 1 or -1, as these are the only possible actions on the group $\mathbb{Z}$. To determine the sign we just note that $\lambda\colon F \times I \to \mathbb{C}^n - Q$ given by $\lambda(\mathbf{z}, t) = e^{\pi i t} \cdot \mathbf{z}$ provides a lift for the necessary square. Putting $t = 1$ we obtain the map $R\colon S^{n-1}_{\mathbb{C}} \to S^{n-1}_{\mathbb{C}}$ given by $(z_1, \dots, z_n) \mapsto (-z_1, -z_2, \dots, -z_n)$. The commutative square

$$\begin{tikzcd} S^{n-1}_{\mathbb{C}} \arrow[r, "R"] & S^{n-1}_{\mathbb{C}} \\ S^{n-1} \arrow[r, "R|S^{n-1}"] \arrow[u, tail, "\sim"] & S^{n-1} \arrow[u, tail, "\sim"'] \end{tikzcd}$$

allows us to identify the map R^* on $H^{n-1}(S^{n-1}_{\mathbb{C}})$ with multiplication by $(-1)^n$. $\square$

It is possible to refine the monodromy action to a slightly different invariant. To see this, let $E \to B$ be a fiber bundle, $b \in B$ be a base point, and let F be the fiber over b. Let $\gamma\colon I \to B$ be a loop based at b and consider again the lifting square

$$\begin{tikzcd} F \times \{0\} \arrow[r] \arrow[d, tail, "\sim"'] & E \arrow[d] \\ F \times I \arrow[r] \arrow[ur, dotted, "\lambda"] & B \end{tikzcd}$$

where the bottom map is the composite $F \times I \to I \xrightarrow{\sigma} B$. Then λ maps $F \times \partial I$ into F, so we can write down the following composite:

$$H_k(F) \xrightarrow{(-)\times i} H_{k+1}(F \times I, F \times \partial I) \xrightarrow{\lambda_*} H_{k+1}(E, F),$$

where $i \in H_1(I, \partial I)$ is the standard generator. The composite is called the **extension map**, or the **variation map**, associated to the path γ. We will denote it by Var_γ. The formula $\partial(x \times i) = (\partial x) \times i + (-1)^{|x|} x \times \partial i$ immediately yields the following connection between variation and monodromy, for $x \in H_k(F)$:

$$\partial[\mathrm{Var}_\gamma(x)] = (-1)^k \cdot [\gamma.x - x]. \tag{2.7}$$

There is also a relative version of the variation. Suppose that $E' \subseteq E$ is a sub-fiber bundle over B, and let $F' \subseteq F$ be the fiber over b. Assume that ???? This time consider the composite

$$\begin{tikzcd} H_{n-1}(F, \partial F') \arrow[r, "(-)\times i"] & H_n(F \times I, (F' \times I) \cup (F \times \partial I)) \arrow[r, "\lambda_*"] & H_n(E, F \cup E') \\ & & H_n(E, F). \arrow[u, "\cong"'] \end{tikzcd}$$

This is called the **relative variation map**, and will be denoted $\mathrm{Var}_\sigma^{rel}$. There is no simple formula connecting the relative variation to monodromy, but the are some evident commutative diagrams connecting the relative variation to the variation.

Now return to the bounded sum-of-squares map $g\colon E \to B$. Let

$$E' = \Big\{(z_1, \dots, z_n) \in \mathbb{C}^n \,\Big|\, \sum_j |z_j|^2 = \epsilon^2 \text{ and } |z_1^2 + \cdots + z_n^2| \le \rho\Big\}$$

and let

$$E'_{sh} = \Big\{(z_1, \dots, z_n) \in \mathbb{C}^n \,\Big|\, \frac{\epsilon^2}{2} \le \sum_j |z_j|^2 \le \epsilon^2 \text{ and } |z_1^2 + \cdots + z_n^2| \le \rho\Big\}.$$

We think of E'_{sh} as a small shell around E'; note that this shell deformation retracts down onto E'. The map $E'_{sh} \to B$ is a fiber bundle, as the map g has no crtical points inside of E'_{sh}. Since B is contractible, we have that $E'_{sh} \cong F'_{sh} \times B$. This homeomorphism can be chosen to carry E' to $(\partial F) \times B$. As a consequence, there is a deformation retraction $r\colon E' \to (\partial F)$. Later we will construct such a map explicitly.

Let $\sigma\colon I \to B$ be a loop. Consider now the lifting square

$$\begin{tikzcd} (F \times \{0\}) \cup (\partial F \times I) \arrow[r] \arrow[d, hook, "\sim"'] & E \arrow[d] \\ F \times I \arrow[r] \arrow[ur, dotted, "\lambda"] & B \end{tikzcd}$$

where the bottom map is the composite $F \times I \to I \xrightarrow{\sigma} B$. Note that $\partial(F \times I) = [(\partial F) \times I] \cup [F \times \partial I]$, and λ carries the first piece into E' and the second piece into F. So λ maps $\partial(F \times I)$ into $F \cup E'$. We can therefore write down the following composite:

$$H_{n-1}(F, \partial F) \xrightarrow{(-)\times i} H_n(F \times I, \partial(F \times I)) \xrightarrow{\lambda_*} H_n(E, F \cup E') \xleftarrow{\cong} H_n(E, F)$$

where i is the canonical generator for $H_1(I, \partial I)$. This composite is called the

PROPOSITION 2.8. *Choose a diagram*

$$\begin{tikzcd} S^{n-1} \arrow[r, "\delta"] \arrow[d, tail] & F \arrow[d, tail] \\ D^n \arrow[r, "\Delta"] & E \end{tikzcd}$$

consisting of a vanishing cycle δ and a corresponding thimble Δ. Let γ be the loop in B given by $t \mapsto \rho e^{2\pi i t}$. Then

$$\mathrm{Var}_\gamma(x) = (-1)^{\binom{n}{2}} \langle x, \delta \rangle \cdot \Delta$$

where $\langle -, - \rangle$ denotes the intersection product in $H_*(F)$ and x is any element of $H_{n-1}(F, \partial F) \cong \mathbb{Z}$.

The proof of this result is not so easy to understand. Note that $H_n(E, F)$ is a copy of $\mathbb{Z}$ generated by Δ, and so $\mathrm{Var}_\gamma(x)$ will necessarily be a multiple of Δ. To prove the result one only has to calculate the coefficient, which for a chosen generator of $H_{n-1}(F, \partial F)$ comes down to just determining a sign. It is a simple matter to produce the lifting λ, and an analysis of this lift determines the sign, but what the lift produces is really an element of $H_n(E, F \cup E')$—whereas Δ is an element of $H_n(E, F)$. Although the groups are isomorphic, it takes some effort to see how a given element of the former determines an element in the latter.

Because of the delicate nature of the proof, we postpone it until Section ????, by which point the reader has a better idea why Proposition 2.8 is important to us.

3. Lefschetz pencils

If V is a complex vector space write $\mathbb{P}(V)$ for the projective space of complex lines in V. In this notation $\mathbb{C}P^n = \mathbb{P}(\mathbb{C}^{n+1})$. If the coordinates on $\mathbb{C}^{n+1}$ are $z_0, z_1, \ldots, z_n$, then a **hyperplane** in $\mathbb{C}P^n$ is the vanishing set for a linear form $a_0 z_0 + \cdots + a_n z_n$, with $a_i \in \mathbb{C}$. Scaling the form does not change the hyperplane, so the hyperplanes in $\mathbb{C}P^n$ are parameterized by the points in the space

$$\widehat{\mathbb{C}P^n} = \mathbb{P}\big(\mathrm{Hom}(\mathbb{C}^{n+1}, \mathbb{C})\big).$$

A **pencil of hyperplanes** is defined to be a linear embedding $H \colon \mathbb{C}P^1 \hookrightarrow \widehat{\mathbb{C}P^n}$, and we use the notation

$$(t \in \mathbb{C}P^1) \mapsto (H_t \subseteq \mathbb{C}P^n).$$

The map H comes from an associated linear embedding $\tilde{H} \colon \mathbb{C}^2 \hookrightarrow \mathrm{Hom}(\mathbb{C}^{n+1}, \mathbb{C})$, and our pencil sends $t = [x : y]$ to $H_t = \mathbb{P}(\ker \tilde{H}(x, y))$. Let f_1 and f_2 be the image under $\tilde{H}$ of the standard basis elements, so that we have $H_{[x:y]} = \mathbb{P}(\ker(x f_1 + y f_2))$. Write

$$W = (\ker f_1) \cap (\ker f_2),$$

which is an $(n-1)$-dimensional subspace of $\mathbb{C}^{n+1}$. Note that $\mathbb{P}(W)$ is contained in H_t for every t. The space $A = \mathbb{P}(W)$ is called the **axis** of the pencil.

The following facts are easy and left to the reader:

(1) Every point in $\mathbb{C}P^n$ is contained in some H_t;
(2) Every point in $\mathbb{C}P^n \backslash A$ is contained in a unique H_t;
(3) If $s \neq t$ then $H_s \cap H_t = A$.

If $X \hookrightarrow \mathbb{C}P^n$ is a subvariety then we would like to define a map $X \to \mathbb{C}P^1$ by sending $x \in X$ to the unique t such that $x \in H_t$. The fibers of such a map would be the intersections $X \cap H_t$. The trouble, of course, is that if $x \in A$ then the map will not be well-defined: because then x belongs to H_t for *every* t! In the case $n = 2$ the space A was just a point, and we could move the pencil so that A did not intersect X. But in higher dimensions this does not work: the dimension of A is $n-2$, and so the intersection $A \cap X$ will be no more than codimension 2 in X. For $\dim X > 1$ this means $A \cap X$ will never be empty, no matter how we choose our pencil.

The way around this problem is to replace X by another space, one obtained by blowing-up A. Define

$$Y = \{(x,t) \,|\, x \in X \text{ and } t \in \mathbb{C}P^1 \text{ and } x \in H_t\} \subseteq X \times \mathbb{C}P^1.$$

Let $p\colon Y \to X$ be the projection onto the first factor, and $f\colon Y \to \mathbb{C}P^1$ be projection onto the second. Let $X' = X \cap A$ and $Y' = p^{-1}(X')$. Note that

(1) $Y \backslash Y' \to X \backslash X'$ is a homeomorphsm;
(2) $Y' = p^{-1}(X') = X' \times \mathbb{C}P^1$;
(3) For each $t \in \mathbb{C}P^1$, the fiber $f^{-1}(t)$ is homeomorphic to $X \cap H_t$.

In this way we have obtained a map $f\colon Y \to \mathbb{C}P^1$ whose fibers are the hyperplane sections of X. This will be our primary object of study.

THEOREM 3.1. *One can choose the embedding $S \hookrightarrow \widehat{\mathbb{C}P^n}$ in such a way that*

(a) The axis A of the pencil intersects X transversally;
(b) Y is smooth;
(c) The map $f\colon Y \to \mathbb{C}P^1$ has finitely-many criticial points, with at most one in each fiber;
(d) Every critical point of f is nondegenerate, in the sense that the determinant of the Hessian matrix of f at this point is nonzero.

A pencil $S \hookrightarrow \widehat{\mathbb{C}P^n}$ with the properties listed in the above theorem is called a **Lefschetz pencil** for X. We will in fact see that such pencils are generic—that is, among the collection of all hyperplane pencils the ones that are not Lefschetz constitute a set of measure zero. This will be explained in Section ???? below. For the moment we wish to accept the existence of Lefschetz pencils and investigate what the consequences are.

Our approach from this point will be to

(i) Study the fibers of the map $f\colon Y \to \mathbb{C}P^1$, and in particular study how the generic fiber degenerates into the singular fibers;
(ii) Compare the homotopy type of the fibers (e.g., the cohomology groups) to the homotopy type of the total space Y;
(iii) Compare the homotopy type of Y to that of X.

In this way we relate the homotopical properties of X to those of the hyperplane sections $X \cap H_t$ which form the fibers of the map f.

To begin with, separate $\mathbb{C}P^1 \cong S^2$ into two disks D_+ and D_-, intersecting in a circle, in such a way that the critical values of f all lie in the interior of D_+. As there are only finitely many critical values, this is certainly possible. Choose a basepoint b in $D_+ \cap D_-$. Define

$$Y_+ = f^{-1}(D_+), \quad Y_- = f^{-1}(D_-), \quad Y_0 = f^{-1}(D_+ \cap D_-), \quad Y_b = f^{-1}(b).$$

Finally, let $n = \dim X$.

As $f: Y_- \to D_-$ has no critical points, it is a fiber bundle—and therefore a trivial fiber bundle since D_- is contractible. So $Y_- \cong D_- \times Y_b$, and $Y_0 \cong S^1 \times Y_b$. We have arranged things so that all the 'action' is concentrated inside of Y_+. The paper [**La**] identifies the main lemma of Lefschetz theory to be the following:

LEMMA 3.2 (Main Lemma). *$H_i(Y_+, Y_b) = 0$ for $i \neq n$ and $H_n(Y_+, Y_b) \cong \mathbb{Z}^r$ where r is the number of critical points of f. Even more, the inclusion $Y_b \hookrightarrow Y_+$ is an ??-equivalence.*

SKETCH OF PROOF. The idea is simple. Use a Mayer-Vietoris argument to reduce the homology calculation to what happens around each critical point. But locally around the critical points we are looking at the bounded sum-of-squares map, which has contractible total space and generic fibers homotopy equivalent to S^{n-1}. So the relative homology group $H_n(U, F)$ around each critical point is isomorphic to $\mathbb{Z}$, and $H_i(U, F) = 0$ for all other i's. Mayer-Vietoris gives that $H_i(Y_+, Y_b)$ is the direct sum of all such $H_i(U, F)$'s, one for each critical point.

A detailed version of this argument can be found in [**La**]. □

The Weak Lefschetz theorem is almost an immediate corollary of the Main Lemma. It is proved by induction, with the induction step being the following:

THEOREM 3.3. *Given a Lefschetz pencil for a dimension n variety $X \hookrightarrow \mathbb{C}P^N$, there is an isomorphism*

$$H_q(X, X_b) \cong H_{q-2}(X_b, X')$$

for $q < n$.

PROOF. Recall that $Y' = p^{-1}(X') \cong X' \times \mathbb{C}P^1$ and we have a homotopy pushout diagram

$$\begin{tikzcd} Y' \arrow[r, tail] \arrow[d] & Y \arrow[d] \\ X' \arrow[r, tail] & X. \end{tikzcd} \tag{3.4}$$

Consider the cofiber sequence

$$(Y_+ \cup Y')/(Y_b \cup Y') \hookrightarrow Y/(Y_b \cup Y') \to Y/(Y_+ \cup Y'). \tag{3.5}$$

The first term is just Y_+/Y_b. The homotopy pushout diagram (3.4) gives that $Y/Y' \to X/X'$ is a weak equivalence, and it then follows that

$$Y/(Y_b \cup Y') \to X/X_b$$

is also a weak equivalence (recall that $X_b \supseteq X'$).

Also, there is a homotopy pushout diagram

$$\begin{tikzcd} Y_0 \arrow[r] \arrow[d] & Y_+ \arrow[d] \\ Y_- \arrow[r] & Y \end{tikzcd}$$

which gives that $Y/Y_+ \simeq Y_-/Y_0$. It likewise follows that $Y/(Y_+ \cup Y') \simeq Y_-/(Y_0 \cup Y')$. But recall $Y_- \cong D_- \times X_b$, and the isomorphism may be chosen to restrict to

$Y_0 \cong (\partial D_-) \times X_b$ and $Y' \cong D_- \times X'$. We then have

$$Y/(Y_+ \cup Y') \simeq (D_- \times X_b)/[(\partial D_- \times X_b) \cup (D_- \times X')] \cong (D_-/\partial D_-) \wedge (X_b/X') \cong S^2 \wedge (X_b/X').$$

We have now proven that (3.5) may be rewritten (in the homotopy category) as a cofiber sequence

$$Y_+/Y_b \to X/X_b \to \Sigma^2(X_b/X').$$

The Main Lemma gives us that Y_+/Y_b is $(n-1)$-connected, and so the homology isomorphisms in the statement of the theorem follow immediately. $\square$

Corollary 3.6 (Weak Lefschetz Theorem). *If $X \hookrightarrow \mathbb{C}P^N$ has dimension d and $H \hookrightarrow \mathbb{C}P^N$ is a generic hyperplane, then $H_q(X, X \cap H) = 0$ for $q < d$. That is, $H_q(X \cap H) \to H_q(X)$ is an isomorphism for $q < d-1$ and an epimorphism for $q = d-1$.*

Proof. We do this by an induction on the dimension of X. If the dimension is zero there is nothing to prove.

If X has dimension d, then we choose a Lefschetz pencil for X and by Theorem 3.3 we have $H_q(X, X_b) \cong H_{q-2}(X_b, X')$ for $q < d$. But X_b is a $(d-1)$-dimensional variety and X' is a generic hyperplane section of X_b, hence by induction $H_{q-2}(X_b, X') = 0$ since $q - 2 < d - 1$. $\square$

4. The Picard-Lefschetz formulas

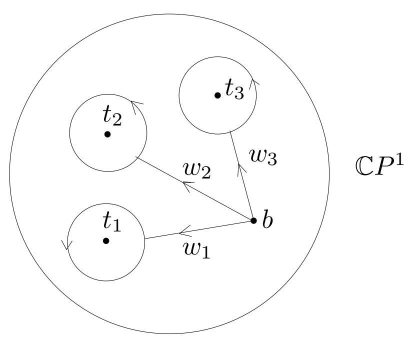
Let us return now to our Lefschetz fibration $f\colon Y \to \mathbb{C}P^1$, with critical values $t_1, \dots, t_k \in \mathbb{C}P^1$. Then

$$Y - f^{-1}(\{t_1, \dots, t_k\}) \to \mathbb{C}P^1 - \{t_1, \dots, t_k\}$$

is a fibration. Pick a $b \in \mathbb{C}P^1 - \{t_1, \dots, t_k\}$ and let Y_b be the fiber over b. We will consider the monodromy action of $\pi_1(\mathbb{C}P^1 - \{t_1, \dots, t_k\}, b)$ on $H^*(Y_b)$ and on $H_*(Y_b)$.

For convenience let $Y^* = Y - f^{-1}(\{t_1, \dots, t_k\})$ and $S^* = \mathbb{C}P^1 - \{t_1, \dots, t_k\}$.

For each i, choose a loop w_i in S^* that is based at b, moves very close to t_i via a simple path, loops once around t_i counterclockwise, and returns to b via the same simple path. What we mean is shown in the following picture:



In fact we will need several pieces of notation that go with this picture. Let D_i be the disk around t_i, let l_i be the path from b to this disk, let s_i be the terminal point of l_i (where it reaches the boundary of the disk), and let γ_i be the counterclockwise loop starting at s_i and moving around the boundary of D_i. Note that

$$w_i = l_i^{-1}\gamma_i l_i.$$

Finally, let $Y_i = f^{-1}(D_i)$, let $F_i = f^{-1}(s_i)$, and let $L_i = f^{-1}(l_i)$.

Provided that the t_i's are appropriately ordered, $\pi_1(S^*, b)$ may be identified with the quotient

$$F(w_1, \dots, w_k)/\langle w_1 w_2 \cdots w_k\rangle_N,$$

where $F(-)$ denotes the free group and $\langle - \rangle_N$ is the normal subgroup generated by the given element. Note that this group is just a free group on $k-1$ generators.

Note that $Y^* \to S^*$ is a fiber bundle and so it is necessarily trivial over l_i. Therefore the inclusions $Y_b \hookrightarrow L_i$ and $F_i \hookrightarrow L_i$ are homotopy equivalences. Consider the following maps of homology groups, all induced by the inclusions:

$$\begin{tikzcd}
H_{n-1}(T_i, F_i) \arrow[r] & H_{n-1}(Y_i, F_i) \arrow[r] & H_{n-1}(Y_+, F_i) \arrow[r, "\cong"] & H_{n-1}(Y_+, L_i) \\
 & & & H_{n-1}(Y_+, Y_b). \arrow[u, "\cong"']
\end{tikzcd}$$

Recall that $H_{n-1}(T_i, F_i) \cong \mathbb{Z}$. A choice of generator gives, via the above maps, an element $\delta_i \in H_{n-1}(Y_+, Y_b)$. Such a homology class will again be called a **vanishing cycle** for the map $f\colon Y \to \mathbb{C}P^1$, just as was done in the local case.

There are three easy facts about the monodromy action of $\pi_1(S^*)$ on $H_*(Y_b)$:

(1) The action on $H_k(Y_b)$ is trivial if $k \neq n$;
(2) $w_i.\delta_i = (-1)^n \delta_i$;
(3) If $\langle x, \delta_i\rangle = 0$ then $w_i.x = x$.

We now explain these three points. First note that we can replace $\mathbb{C}P^1$ by D_+ and Y by Y_+. Write $Y_+^* = Y_+ \cap Y^*$ and $D_+^* = D_+ \cap S^*$. The key point for the rest of the argument is that if $j\colon Y_b \hookrightarrow Y_+^*$ is the inclusion then $j_*(w_i.x) = j_*(x)$ for all $x \in H_*(Y_b)$. This is a basic fact about monodromy that we will explain below. By the Main Lemma (3.2) the map $H_k(Y_b) \to H_k(Y_+)$ is injective for $k \neq n-1$, and consequently the same is true for j_*. So it follows that $w_i.x = x$ when $x \in H_k(Y_b)$ and $k \neq n-1$.

To explain the key point about the monodromy action from the last paragraph, recall that the action of w_i comes about from a lifting square

$$\begin{tikzcd}
Y_b \times \{0\} \arrow[r] \arrow[d] & Y_+^* \arrow[d] \\
Y_b \times I \arrow[r] \arrow[ur, "\lambda"] & D_+^*
\end{tikzcd}$$

where the bottom map is the composite $Y_b \times I \to I \xrightarrow{w_i} D_+^*$. More precisely, $\lambda_1 = \lambda|_{Y_b \times \{1\}}$ gives a map $Y_b \to Y_b$ and the induced map on homology is the monodromy action of w_i. But if we look in Y_+^* instead of in Y_b (i.e., if we compose with the inclusion $j\colon Y_b \hookrightarrow Y_+^*$) then λ itself is precisely a homotopy between λ_1 and the identity. In other words, $j \circ \lambda_1$ is homotopic to j.

Fact (2) is a direct consequence of ?????.

Fact (3) uses a geometric consequence of Lefschetz Duality. Namely, Proposition 4.1 (Lefschetz Moving Lemma). *Suppose that M and N are compact, n-dimensional manifolds-with-boundary, and that they have the same boundary: $\partial M = \partial N = A$. Let $X = M \cup_A N$, which is a compact n-manifold. Assume that X is orientable. Let $x \in H_k(X)$. If $\langle x, i_*(y)\rangle = 0$ for all $y \in H_{n-k}(N)$, then x is in the image of $H_k(M) \to H_k(X)$.*

In the situation of the above result result, note that if x is in the image of $H_k(M)$ then it is also in the image of $H_k(M \backslash \partial M)$, because by choosing a collaring of the boundary we can deformation retract M into $M \backslash \partial M$. So any homology class in the image of $H_k(M)$ will have trivial intersection with classes in the image of $H_k(N)$. This works with any coefficients. The Lefschetz Moving Lemma gives a converse to this in the case of *rational* coefficients: if a cycle in X has trivial intersection product with all cycles lying in N, then it can be moved into M.

Proof of Proposition 4.1. First consider the sequence

$$H_k(M) \xrightarrow{j_*} H_k(X) \xrightarrow{l_*} H_k(X, M),$$

which is exact in the middle, together with the excision isomorphism $H_k(N, \partial N) \xrightarrow{i_*} H_k(X, M)$. Then $x \in H_k(X)$ is in the image of j_* if and only if $(i_*)^{-1}(l_*(x))$ is zero. But for $u \in H_*(N)$, intersection theory gives us that

$$\langle (i_*)^{-1}(l_*(x)), u\rangle_N = \langle x, i_*(u)\rangle_X,$$

and we have assumed that this vanishes for every u. Lefschetz Duality says that the intersection pairing $H_k(N, \partial N) \otimes H_{n-k}(N) \to \mathbb{Q}$ is nondegenerate, so the above implies that $(i_*)^{-1}(l_*(x)) = 0$. This means that $l_*(x) = 0$, and so x is in the image of j_*. $\square$

Now let us apply the Lefschetz Moving Lemma to our monodromy calculation. Let $x \in H_{n-1}(Y_b)$ and assume that $\langle x, \delta_i\rangle = 0$. By applying the monodromy for l_i, we may as well replace Y_b with F_i. Let B be a small closed ball around the critical point of f, and note that $F_i \backslash \operatorname{int} B$ and $F_i \cap B$ are two n-dimensional manifolds with boundary, whose union is F_i. Since $H_{n-1}(F_i \cap B) \cong \mathbb{Z}$ with generator δ_i, our assumption that $\langle x, \delta_i\rangle = 0$ together with the Lefschetz Moving Lemma implies that x is in the image of $H_{n-1}(F_i \backslash \operatorname{int} B)$. But $Y_i \backslash \operatorname{int} B \to D_i$ is a fiber bundle (we have removed the single criticial point from Y_i, which lies in B), and for this bundle the monodromy action of w_i is trivial because w_i vanishes inside of $\pi_1(D_i, b)$. It follows by naturality that the monodromy action of w_i, on the class x, with respect to the bundle $Y_i \to D_i^*$ is also trivial. This completes the argument.

Facts (1)–(3) are almost enough to determine the entire monodromy action of $\pi_1(S^*, b)$ on $H_*(Y_b)$. When n is odd they *are* enough, but when n is even one needs to work harder. The result we are aiming for, which gives a complete description of the monodromy action in all cases, is the following:

Theorem 4.2 (Picard-Lefschetz). *$\pi_1(S^*, b)$ acts trivially on $H_q(Y_b)$ for $q \neq n-1$. For $x \in H_{n-1}(Y_b)$ the action is given by*

$$w_i.x = x + (-1)^{\binom{n+1}{2}} \langle x, \delta_i\rangle \delta_i.$$

The displayed formula in this theorem is called the **Picard-Lefschetz formula**. Before proving it we need one small lemma:

LEMMA 4.3. *Let $\delta_i \in H_{n-1}(Y_b)$ be a vanishing cycle. Then*

$$\langle \delta_i, \delta_i \rangle = \begin{cases} 0 & \text{if } n-1 \text{ is odd} \\ 2 \cdot (-1)^{\frac{n-1}{2}} & \text{if } n-1 \text{ is even.} \end{cases}$$

PROOF. When $n-1$ is odd there is nothing to prove, as the intersection pairing is skew-symmetric on odd-dimensional homology groups. (The vanishing could also be proven by the same argument we are about to give for the even case.)

As a general remark, it suffices to prove the theorem with Y_b replaced with any other non-critical fiber; so in particular, we can move b as close to the critical value t_i as we would like. Choose a neighborhood around the critical point c_i in which the map f may be written as the sum-of-squares map. Within this neighborhood Y_b is homeomorphic to $D(TS^{n-1})$, the disk bundle of the tangent bundle for S^{n-1}, and under this homeomorphism δ_i comes from the map $S^{n-1} \to D(TS^{n-1})$ given by the zero-section. Since the intersection product is local, we can replace Y_b with $D(TS^{n-1})$ and δ_i with the zero-section. But then the intersection product gives the Euler characteristic of S^{n-1}, by a classical calculation. So the intersection product is 2 when $n-1$ is even (and zero when $n-1$ is odd). There is one last point to be made, though, which is that the natural orientation of $D(TS^{n-1})$ does not coincide with the natural orientation on Y_b (the one coming from the complex structure). Here one needs to recall the homeomorphism between the complex manifold $z_1^2+\cdots+z_n^2 = 1$, $|(z_1,\ldots,z_n)| \leq \epsilon$ and $D(TS^{n-1})$, which sends $(z_1,\ldots,z_n)$ to the tuple $(\frac{u}{|u|}, v)$ where $z_j = u_j + iv_j$ and $u = (u_1,\ldots,u_n)$, $v = (v_1,\ldots,v_n)$. Near the point $z = (1,0,0,\ldots,0)$, the orientation on our complex manifold is represented by the coordinate system $u_2, v_2, \ldots, u_n, v_n$, whereas the orientation on the disk bundle near this point is represented by $u_2,\ldots,u_n,v_2,\ldots,v_n$. These orientations differ by $(-1)^{\binom{n-1}{2}}$. But for $n = 2k+1$ one has

$$\tbinom{n-1}{2} = k(2k-1) \equiv k = \tfrac{n-1}{2} \pmod 2.$$

□

PROOF OF THEOREM 4.2. The statement about the action on $H_q(Y_b)$ for $q \neq n-1$ was explained above in point (1). For the action on $H_{n-1}(Y_b)$, the case where n is odd is easy and the case where n is even is hard. When n is odd we know that $\langle \delta_i, \delta_i \rangle \neq 0$, and so the space $H_{n-1}(Y_b)$ is the direct sum of $\mathbb{Q}.\delta_i$ and $(\mathbb{Q}.\delta_i)^\perp$ (the point is that the intersection form is nondegenerate and these two subspaces do not intersect) Since we know the action $w_i.(-)$ on each summand by points (2) and (3) above, we know the action everywhere. It is just a matter of checking that the Picard-Lefschetz formula conforms with the formulas in (2) and (3), which is easy (using Lemma 4.3 for (2)).

In the case when n is even, $\langle \delta_i, \delta_i \rangle = 0$ and so $\delta_i \in (\mathbb{Q}.\delta_i)^\perp$. So formulas (2) and (3) do not suffice to determine the action here. What remains is to determine $w_i.c$ for a $c \in H_{n-1}(Y_b)$ with $\langle c, \delta_i \rangle = 1$. This is where things get tricky. The key is to calculate the variation, and then to use the variation to get the monodromy.

Let ???????. Consider the following diagram:

$$\begin{tikzcd} H_{n-1}(Y_b) \arrow[r] \arrow[dr, "\mathrm{Var}_{\gamma_i}"'] & H_{n-1}(Y_b, Y_b \setminus \operatorname{int} F) \arrow[d, "\mathrm{Var}^{rel}_{\gamma_i}"] & H_{n-1}(F, \partial F) \arrow[l, "\cong"'] \arrow[d, "\mathrm{Var}^{rel}_{\gamma_i}"] \\ & H_n(Y_+, Y_b) & H_n(T, F) \arrow[l] \end{tikzcd}$$

Let $C \in H_{n-1}(F, \partial F)$ be the image of c under the top composite. Then $\langle C, \delta_i \rangle = 1$, and in fact C is completely determined by this property. The right-most vertical map has been calculated previously, as this is just a local matter. By Proposition 2.8 we know that

$$\mathrm{Var}^{rel}_{\gamma_i}(C) = (-1)^{\binom{n}{2}} \cdot \Delta_i.$$

By naturality we obtain $\mathrm{Var}_{\gamma_i}(c) = (-1)^{\binom{n}{2}} i_*(\Delta_i)$. Finally, recall the formula

$$\partial \,\mathrm{Var}_{\gamma_i}(x) = (-1)^n [\gamma_i.x - x]$$

from (2.7). Since $\partial(\Delta_i) = \delta_i$, by combining the above formulas we now get

$$\gamma_i.x = x + (-1)^{\binom{n}{2}+n}\delta_i = x + (-1)^{\binom{n+1}{2}}\delta_i.$$

To complete the proof, just note that $H_{n-1}(Y_b)$ is the direct sum of $\mathbb{Q}.c$ and $(\mathbb{Q}.\delta_i)^\perp$. We just verified the Picard-Lefschetz formula on the first piece, and on the second piece it was already verified by point (3) above. $\square$

Let $V \subseteq H_{n-1}(Y_b)$ be the $\mathbb{Q}$-subspace generated by the vanishing cycles $\delta_1, \ldots, \delta_k$. Define $I \subseteq H_{n-1}(Y_b)$ to be the set of elements fixed by the monodromy action; that is, $I = [H_{n-1}(Y^b)]^{\pi_1(S^*)}$. Then we can state the following

Corollary 4.4. $I = V^\perp$.

Proof. This follows directly from the Picard-Lefschetz formulas. $\square$

4.5. Consequences of the Picard-Lefschetz formulas.

Proposition 4.6. *For any two vanishing cycles δ_i and δ_j, there exists a $g \in G$ such that $g.\delta_i = \pm\delta_j$.*

????

Theorem 4.7. *The following conditions are equivalent:*

(1) The restriction of the intersection form $\langle -, - \rangle$ to $V \subseteq H_{n-1}(Y_b)$ is nondegenerate;
(2) Either $V = 0$ or V is a non-trivial simple G-module (i.e., V is simple but it is not the one-dimensional trivial module);
(3) $H_{n-1}(Y_b)$ is a semi-simple G-module.

Proof. Suppose $\langle -, - \rangle$ is nondegenerate on V. Let $W \subseteq V$ be a nonzero G-invariant submodule, and let $x \in W - \{0\}$. By the nondegeneracy of the form, there exists a δ_i such that $\langle x, \delta_i \rangle \neq 0$. By the Picard-Lefschetz formula we have $(w_i).x = x \pm \langle x, \delta_i \rangle \delta_i$. Since x and $(w_i).x$ both belong to W, it follows that $\delta_i \in W$ as well. Then by Proposition 4.6 all the other δ_j's also belong to W, hence $W = V$. So we have proven that V is simple. If $n-1$ is odd then the form $\langle -, - \rangle$ is alternating, hence non-degeneracy guarantees that V is even-dimensional (and in particular, is not one-dimensional). If $n-1$ is even then the Picard-Lefschetz formula gives

$$(w_i).\delta_i = -\delta_i,$$

hence the G-action on V is not trivial. We have therefore proven that (1) implies (2).

Note that $V \cap V^\perp$ is a G-invariant submodule of V. If (2) holds, then either $V \cap V^\perp = 0$ (in which case the form $\langle -, - \rangle$ is nondegenerate on V) or else $V \cap V^\perp = V$. But the latter yields $V \subseteq V^\perp$, yet G acts trivially on $V^\perp$ by Picard-Lefschetz and nontrivially on V by assumption. This proves (2) implies (1).

Assumption (1) yields that $V \cap V^\perp = 0$, hence $H_{n-1}(Y_b)$ decomposes as $V \oplus V^\perp$. The group G acts trivially on $V^\perp$, and by (1) $\Rightarrow$ (2) we know that V is simple. So $H_{n-1}(Y_b)$ is semi-simple.

Finally, suppose (3) is true and let $v \in V^\perp$ be nonzero. Then $g.v = v$ for all $g \in G$. If L is the subspace spanned by v, it follows that $L^\perp$ is G-invariant: for if $\langle v, x\rangle = 0$ then $\langle v, g.x\rangle = \langle g.v, g.x\rangle = \langle v, x\rangle = 0$ as well. Since $H_{n-1}(Y_b)$ is assumed to be semi-simple, it follows that there is a G-submodule M such that $H_{n-1}(Y_b) = L^\perp \oplus M$. By nondegeneracy of the form on $H_{n-1}(Y_b)$, the module M is one-dimensional. But if $m \in M - \{0\}$ then $\langle v, m\rangle = \langle g.v, g.m\rangle = \langle v, g.m\rangle$, and so $\langle v, g.m - m\rangle = 0$. Therefore $g.m - m \in L^\perp \cap M$, hence $g.m = m$. As this holds for all g, we have that M is the trivial G-module. So M is contained in $V^\perp$.

Now suppose that we also have $v \in V$. Then M is orthogonal to v, hence M is contained in $L^\perp$. But this is a contradiction. So $V \cap V^\perp = 0$, and this proves that (3) implies (1). $\square$

5. Construction of Lefschetz pencils

In this section we develop the machinery needed to establish the existence of Lefschetz pencils. Even if you were willing to accept their existence on faith, this machinery is still needed to prove some of the basic facts about the monodromy action.

5.1. Geometry of the dual variety. Consider the projective space $\mathbb{C}P^n$ and the dual projective space $\widehat{\mathbb{C}P^n}$. Recall that the points of $\widehat{\mathbb{C}P^n}$ are in bijective correspondence with hyperplanes in $\mathbb{C}P^n$. Given a subset $S \subseteq \mathbb{C}P^n$, define $\alpha(S) \subseteq \widehat{\mathbb{C}P^n}$ to be the set of hyperplanes in $\mathbb{C}P^n$ that contain S. That is,

$$\alpha(S) = \{H \in \widehat{\mathbb{C}P^n} \mid H \supseteq S\}.$$

Likewise, given $T \subseteq \widehat{\mathbb{C}P^n}$ define

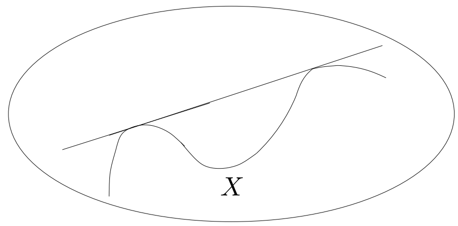
$$\beta(T) = \bigcap_{H \in T} H.$$

Note that if $S \subseteq S'$ then $\alpha(S) \supseteq \alpha(S')$, and similarly for β. That is, both α and β reverse the order of subsets.

Linear algebra shows that if S is a linear subspace of $\mathbb{C}P^n$ then $\beta(\alpha(S)) = S$, and if T is a linear subspace of $\widehat{\mathbb{C}P^n}$ then $\alpha(\beta(T)) = T$. The maps α and β give a bijective correspondence between the linear k-dimensional subspaces of $\mathbb{C}P^n$ and the linear $(n-k-1)$-dimensional subspaces of $\widehat{\mathbb{C}P^n}$.

Let $X \subseteq \mathbb{C}P^n$ be a smooth hypersurface. Then we can define a map of spaces $\Phi\colon X \to \widehat{\mathbb{C}P^n}$ by letting $\Phi(x)$ be the tangent hyperplane to X at the point x. The map Φ is algebraic; if X is defined by the homogeneous polynomial equation $g = 0$ then in appropriate coordinates $\Phi(x)$ is just $(\nabla g)(x)$ (more precisely, $\Phi(x)$ is the projective space of the kernel of the linear map $\mathbb{C}^{n+1} \to \mathbb{C}$ whose associated matrix is $(\nabla g)(x)$). The image of Φ will be denoted $D(X)$, and called the **dual variety** of X. Notice that $D(X)$ is simply the collection of hyperplanes that are tangent to X at some point.

A general fact about $D(X)$ is that it will be $(n-1)$-dimensional at every point—one can imagine moving a tangent hyperplane in any direction along X, and $\dim X = n - 1$. But unlike X, the dual variety $D(X)$ might be singular. This will happen exactly when there exist hyperplanes that are simultaneously tangent to X at more than one point, as depicted below:



In this case the map Φ is not injective, and so points of X are being identified to make $D(X)$.

Since $D(X)$ is possibly not smooth, we cannot form $D(D(X))$ using our present definition. However, there is a clever generalization of our construction which works for any subvariety of $\mathbb{C}P^n$: it doesn't have to be smooth, and it doesn't have to be a hypersurface. We will now describe this generalization and use it to prove that $D(D(X)) \cong X$. This will justify the name "dual variety".

Let $X \subseteq \mathbb{C}P^n$ be a closed, algebraic subvariety. When $x \in X$ is a smooth point, let us say that a hyperplane is "tangent to X at x" if it contains the tangent space T_xX. When x is not a smooth point, we will regard a hyperplane as being tangent to X at x if it is the limit of hyperplanes that are tangent to X at smooth points. We then define $D(X) \subseteq \widehat{\mathbb{C}P^n}$ to be the collection of all tangent hyperplanes.

To make the above description completely rigorous we do the following. Let $X_{ns} \subseteq X$ denote the Zariski open set of nonsingular points. Define a subspace $V_X^\circ \subseteq \mathbb{C}P^n \times \widehat{\mathbb{C}P^n}$ by

$$V_X^\circ = \{(x, H) \in \mathbb{C}P^n \times \widehat{\mathbb{C}P^n} \,|\, x \in X_{ns} \text{ and } H \text{ is tangent to } X \text{ at } x\}.$$

Then V_X° is a quasi-projective subvariety of $\mathbb{C}P^n \times \widehat{\mathbb{C}P^n}$. Let V_X be the Zariski closure of V_X°. There are maps $\pi_1 \colon V_X \to \mathbb{C}P^n$ and $\pi_2 \colon V_X \to \widehat{\mathbb{C}P^n}$. We define $D(X) = \pi_2(V_X)$ and call this the **dual variety** of X.

Note the following facts:

(1) When X is smooth, $V_X^\circ = V_X$.
(2) When X is smooth, $\pi_1 \colon V_X \to X$ is a fiber bundle with fiber $\mathbb{C}P^{n-1-\dim X}$ (this is the collection of hyperplanes in $\mathbb{C}P^n$ containing a given tangent space of X). In particular, if X is irreducible then so is V_X.
(3) When X is a smooth hypersurface, $\pi_1 \colon V_X \to X$ is an isomorphism. Note that in this case $\pi_2 \colon V_X \to \widehat{\mathbb{C}P^n}$ is just a map $X \to \widehat{\mathbb{C}P^n}$, and it coincides with the map Φ introduced earlier in this section.

5.2. The basic construction. We are almost ready to describe the construction of Lefschetz pencils. In brief, the idea is to choose a projective line $S \hookrightarrow \widehat{\mathbb{C}P^n}$ that avoids the singular set of $D(X)$ and intersects $D(X)$ transversally. This line S will be our pencil of hyperplanes. We will show to construct the Lefschetz fibration $Y \to S$ and prove that this has the properties we outlined in Theorem 3.1.

From now on assume that X is smooth. Let

$$W = \{(x, H) \in \mathbb{C}P^n \times \widehat{\mathbb{C}P^n} \,|\, x \in H\}.$$

Note that the projection $W \to \mathbb{C}P^n$ is a fiber bundle with fiber $\mathbb{C}P^{n-1}$ (and likewise for the projection $W \to \widehat{\mathbb{C}P^n}$). Let W_X be the pullback of $W \to \mathbb{C}P^n$ to X. That is,

$$W_X = \{(x, H) \in \mathbb{C}P^n \times \widehat{\mathbb{C}P^n} \,|\, x \in X \cap H\}.$$

Note that V_X is a subspace of W_X. We have the following diagram depicting the spaces we are currently considering:

$$\begin{tikzcd}
 & & & \widehat{\mathbb{C}P^n} \\
V_X \arrow[r, tail] \arrow[drr, bend right, two heads, "\mathbb{C}P^{n-1-\dim X}"'] & W_X \arrow[r, tail] \arrow[dr, two heads, "\mathbb{C}P^{n-1}"'] & W \arrow[r, tail] \arrow[ur, two heads, "\mathbb{C}P^{n-1}"] \arrow[dr, two heads, "\mathbb{C}P^{n-1}"'] & \mathbb{C}P^n \times \widehat{\mathbb{C}P^n} \arrow[u, "\pi_2"'] \arrow[d, "\pi_1"] \\
 & & X \arrow[r, tail] & \mathbb{C}P^n
\end{tikzcd}$$

Here the two-headed arrows are fiber bundles, with the fiber indicated next to the arrow.

Now assume that X is both smooth and irreducible. Then W_X is also smooth and irreducible, as $W_X \to X$ is a fiber bundle with fiber $\mathbb{C}P^{n-1}$. We will study in detail the map $\pi_2 \colon W_X \to \widehat{\mathbb{C}P^n}$. Note that the fiber of π_2 over a point $H \in \widehat{\mathbb{C}P^n}$ is the intersection $X \cap H$.

Define Y to be the pullback of $\pi_2 \colon W_X \to \widehat{\mathbb{C}P^n}$ along $S \hookrightarrow \widehat{\mathbb{C}P^n}$:

$$\begin{tikzcd}
Y \arrow[r] \arrow[d] & W_X \arrow[d] \\
S \arrow[r, tail] & \widehat{\mathbb{C}P^n}.
\end{tikzcd}$$

The map $Y \to S$ will be our Lefschetz fibration. Our task, then, is to analyze the critical points of this map.

5.3. Detailed analysis of critical points. It will be important to first analyze the critical points of $W_X \to \widehat{\mathbb{C}P^n}$, which we do below:

Lemma 5.4.

(a) The critical points of $\pi_2 \colon W_X \to \widehat{\mathbb{C}P^n}$ are those pairs (x, H) where H is tangent to X at x. In other words, the critical points of π_2 constitute the subspace V_X.

(b) If (x, H) is a critical point of π_2, then $(D\pi_2)(T_{(x,H)}W_X)$ is precisely $T_H(D(X))$.

Proof. Part (a) is clear geometrically. If (x, H) is a point in W_X where H is not tangent to X, then all small movements of H still result in a nearby intersection with X. This amounts to saying that $(D\pi_2)_{(x,H)}$ is surjective.

However, when H is tangent to X then moving H in the normal direction to X does not result in a nearby intersection. So this "direction" in the tangent space $T_H(\widehat{\mathbb{C}P^n})$ is not in the image of $(D\pi_2)_{(x,H)}$.

Part (b) involves the same ideas as the previous paragraph. The image of $(D\pi_2)_{(x,H)}$ constitutes the "directions of movement" that result in H still intersecting X near x. But for a tangent hyperplane these are precisely the movements where H remains tangent to X. This exactly describes $T_H(D(X))$. $\square$

Proposition 5.5 (Duality Theorem). *Assume that X is smooth and irreducible. Let $x \in \mathbb{C}P^n$, and let $H_x \in \widehat{\mathbb{C}P^n}$ be the dual hyperplane—that is, $H_x = \alpha(\{x\})$. Then $x \in X$ if and only if H_x is tangent to $D(X)$. Equivalently, $D(D(X)) = X$.*

This is in some sense repeating the obvious, but it is nice to think of this result via the following picture:

As the point x moves in $\mathbb{C}P^n$, the associated hyperplane H_x moves around $\widehat{\mathbb{C}P^n}$. The Duality Theorem says that when x moves onto X the hyperplane H_x becomes tangent to $D(X)$, and vice versa.

Proof. Using the evident identification $\widehat{\widehat{\mathbb{C}P^n}} = \mathbb{C}P^n$, the variety $V_{D(X)}$ is just

$$V_{D(X)} = \{(x, H) \in \mathbb{C}P^n \times \widehat{\mathbb{C}P^n} \mid H_x \text{ is tangent to } D(X) \text{ at } H\}.$$

The statement of the proposition amounts to the equality $V_X = V_{D(X)}$. This is because $x \in X$ if and only if there exists an H such that $(x, H) \in V_X$, and likewise H_x is tangent to $D(X)$ if and only if there exists an H such that $(x, H) \in V_{D(X)}$.

We will prove that $V_X \subseteq V_{D(X)}$. The subset in the other direction is nearly the same, but it will be easier to observe that both V_X and $V_{D(X)}$ are closed, irreducible subvarieties, and moreover they have the same dimension. So once we know $V_X \subseteq V_{D(X)}$ we automatically have equality.

Define $U \subseteq V_X$ to be the set of points $(x, H) \in V_X$ such that $H \in D(X)_{ns}$ (in other words, H is tangent to X at the point x and nowhere else). Then U is a nonempty, Zariski open subset of V_X, and therefore $\overline{U} = V_X$. We need only prove $U \subseteq V_{D(X)}$, since then we have $V_X = \overline{U} \subseteq V_{D(X)}$ and we are done.

Let $(x, H) \in U$. In vague terms, the proof goes as follows. The space $T_H[D(X)]$ consists of all "infinitesimal movements" of H that remain tangent to X. It is clear geometrically that this is the same thing as all "infinitesimal movements" of H that intersect X near x. Because H is tangent to X, this space is $(n-1)$ dimensional. But at the same time it clearly contains all movements where H is simply "rotated" about the point x, and these constitute the $(n-1)$-dimensional space H_x. So we have that H_x contains $T_H[D(X)]$, i.e. that H_x is tangent to $D(X)$ at H.

We now phrase the above argument more formally. Note that $\{x\} \times H_x \subseteq W_X$; that is, $(x, J) \in W_X$ for any hyperplane J that contains x. Consider the subsets

$$T_H(H_x) \subseteq (D\pi_2)(T_{(x,H)}W_X) \supseteq (D\pi_2)(T_{(x,H)}V_X) \subseteq T_H[D(X)].$$

The assumption $(x, H) \in U$ gives that the last of these is an equality. The space $T_H(H_x)$ has dimension $n-1$, and since (x, H) is a critical point for $\pi_2 \colon W_X \to \widehat{\mathbb{C}P^n}$ we have that the dimension of $(D\pi_2)(T_{(x,H)}W_X)$ is at most $n-1$. So the first subset is also an equality. We therefore conclude that $T_H(H_x) \supseteq T_H[D(X)]$, which is what we wanted. So $(x, H) \in V_{D(X)}$, we have shown $U \subseteq V_{D(X)}$, and we are done. $\square$

Before proceeding further we need a basic fact from algebraic geometry. Let $b \in \mathbb{C}P^n$, which corresponds to a line $l \subseteq \mathbb{C}^{n+1}$. Choose a complementary subspace W to this line. Then the projective lines in $\mathbb{C}P^n$ that pass through b are in bijective correspondence with the points of $\mathbb{P}(W)$. We denote $\mathcal{L}_b = \mathbb{P}(W)$ and call this the space of projective lines through b.

Lemma 5.6. *Let $Z \subseteq \mathbb{C}P^n$ be a closed, proper subvariety and let $b \in \mathbb{C}P^n \backslash Z$.*

(a) If $\dim Z \leq n-2$ then the set of projective lines through b that do not meet Z is a nonempty, Zariski open subset of $\mathcal{L}_b$.

(b) If $\dim Z = n-1$ then the set of projective lines through b that do not intersect the singular set of Z and also meet Z transversally form a nonempty, Zariski open subset of $\mathcal{L}_b$.

Proof. This is standard. $\square$

Now let us be very specific about our choice of projective line $S \hookrightarrow \widehat{\mathbb{C}P^n}$. Choose any point $b \in \widehat{\mathbb{C}P^n} \backslash D(X)$. Note that this corresponds to a hyperplane $H_b \subseteq \mathbb{C}P^n$ that intersects X transversally. If $\dim D(X) \leq n-2$, let $S \hookrightarrow \mathbb{C}P^n$ be a projective line through b that doesn't intersect $D(X)$. If $\dim D(X) = n-1$ then let S be a projective line through b that avoids the singular set of $D(X)$ and also meets $D(X)$ transversally. As we have seen above, this S determines a pencil of hyperplanes and a corresponding map $f : Y \to S$.

If $t \in S \backslash D(X)$ then H_t intersects X transversally. If $t \in S \cap D(X)$ then H_t is tangent to X in exactly one point x_t (for otherwise H_t would be a singular point of $D(X)$, and S was chosen to avoid such points).

Lemma 5.7. *The critical points of $f : Y \to S$ are the pairs (x, t) such that $t \in S \cap D(X)$ and H_t is tangent to X at x. In particular, the points $S \cap D(X)$ are the critical values of f and each of these values has exactly one critical point in its fiber.*

Proof. By general properties of pullbacks we know that

$$(Df)(T_{(x,H)}Y) = (D\pi_2)(T_{(x,H)}W_X) \cap T_H S.$$

We know by Lemma 5.4(b) that $(D\pi_2)(T_{(x,H)}W_X)$ equals $T_H\mathbb{C}P^n$ when H is not tangent to X at x, and that it equals $T_H[D(X)]$ when H is tangent to X at x. In the former case we have that Df is surjective, and in the latter case it is not surjective because $T_H[D(X)] \cap T_H S$ is 0-dimensional by our choice of S. So the critical points of f are pairs (x, H) where H is tangent to X at x.

If (x, H) and (x', H) are both critical points of f, then H is tangent to X at both x and x'. This would imply that H lies in the singular set of $D(X)$, yet S was chosen to avoid this singular set. So no two critical points of f lie in the same fiber. $\square$

Proposition 5.8 (Existence of Lefschetz pencils). *With $S \subseteq \widehat{\mathbb{C}P^n}$ chosen as above, we have that:*

(a) The axis of the pencil intersects X transversally;

(b) If $\dim D(X) \leq n-2$ the map $f : Y \to S$ has no critical points;

(c) *If* $\dim D(X) = n-1$ *then the map* $f\colon Y \to S$ *has finitely many critical points, the number of such points is the same as the degree of* $D(X)$*, and no two critical points lie in the same fiber.*
(d) *In the situation of part (c), it moreover is true that each critical point of the map* $f\colon Y \to S$ *is nondegenerate.*

PROOF. The axis of the pencil is $A = \cap_{H\in S} H$, and recall that this is just $\beta(S)$. This is a codimension two linear subspace of $\mathbb{C}P^n$. Suppose that A does not intersect X transversally. Then there is a point $x \in A \cap X$ such that $A \subseteq T_x X$. Write $H = T_x X$. Applying α to the inclusions $\{x\} \subseteq A \subseteq H$ yields

$$\alpha(x) \supseteq S \supseteq \alpha(H). \tag{5.9}$$

Since H is tangent to X, $\alpha(H)$ is by definition a point on $D(X)$. So $\alpha(H)$ lies on $S \cap D(X)$. But the Duality Theorem (Proposition 5.5) says that H is tangent to X at x if and only if $\alpha(x)$ is tangent to $D(X)$ at $\alpha(H)$. By (5.9) this implies that S is tangent to $D(X)$ at $\alpha(H)$, which contradicts our choice of S. This proves (a).

Part (b) and (c) follow directly from Lemma 5.7.

For part (d) we have to determine the Hessian for f in the neighborhood of a critical point $(w, H) \in Y$. By the lemma, being a critical point means that H is tangent to X at w.

Recall that $Y \subseteq W_X \subseteq W \subseteq \mathbb{C}P^n \times \widehat{\mathbb{C}P^n}$. Choose projective coordinates $x_0, \ldots, x_n$ on $\mathbb{C}P^n$, and let $y_0, \ldots, y_n$ be the dual coordinates on $\widehat{\mathbb{C}P^n}$. In these coordinates W consists of all pairs $([x_0 : \ldots : x_n], [y_0 : \ldots : y_n])$ such that $\sum x_i y_i = 0$. By choosing the coordinates appropriately, we can assume that $S \hookrightarrow \widehat{\mathbb{C}P^n}$ is the subspace $\{[a : 0 : 0 : \ldots : 0 : b] \mid a, b \in \mathbb{C}\}$, that $w = [1 : 0 : 0 : \ldots : 0]$, and moreover that $H = [0 : 0 : \ldots : 0 : 1]$.

Our first goal is to determine what $\pi_2\colon W_X \to \widehat{\mathbb{C}P^n}$ looks like in local coordinates around the point (x, H). To this end, let $U = \{x \in X \mid x_0 \neq 0\}$ and note that $\pi_1\colon W_X \to X$ is trivial over U. In fact let us use the explicit trivialization

$$\begin{aligned}
&\phi\colon U \times \mathbb{C}P^{n-1} \xrightarrow{\cong} \pi_1^{-1}(U) \\
&(x, [z_1 : \ldots : z_n]) \mapsto \Big(x, \Big[-\sum_{i>0} x_i z_i : x_0 z_1 : x_0 z_2 : \ldots : x_0 z_n\Big]\Big).
\end{aligned}$$

For points in $\mathbb{C}P^n$ near w, we can normalize their projective coordinates so that $x_0 = 1$. For points in $\mathbb{C}P^{n-1}$ near $[0 : 0 : \ldots : 1]$, we can normalize their projective coordinates so that $z_n = 1$. Let $t_1, \ldots, t_d$ be local holomorphic coordinates on X near w: so we have functions $x_i = x_i(t_1, \ldots, t_d)$ and the points $[1 : x_1(t) : \ldots : x_n(t)]$ give a neighborhood of w in X. We find that

$$t_1, \ldots, t_d, z_1, \ldots, z_{n-1} \mapsto \phi([1 : x_1(t) : \ldots : x_n(t)], [z_1 : \ldots : z_{n-1} : 1])$$

give local affine coordinates in W_X near w.

Finally, on $\widehat{\mathbb{C}P^n}$ we get affine coordinates in a neighborhood of H by normalizing so that $y_n = 1$. Using these coordinates we find that $\pi_2\colon W_X \to \widehat{\mathbb{C}P^n}$ has the form

$$t_1, \ldots, t_d, z_1, \ldots, z_{n-1} \mapsto \Big(-\sum_{i>0} x_i(t) z_i, z_1, \ldots, z_{n-1}\Big) = (g(t, z), z_1, \ldots, z_{n-1})$$

where we take the last equality as a definition of the function $g(t, z)$.

Now that we know π_2 in local coordinates, we can compute its Jacobian. We find it has the form

$$\begin{bmatrix} \frac{\partial g}{\partial t_1} & \frac{\partial g}{\partial t_2} & \cdots & \frac{\partial g}{\partial t_d} & * & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Recall that $V_X \subseteq W_X$ is the space of critical points of π_2, i.e. the space of points where the above Jacobian matrix does not have full rank. So in our local coordinate system V_X is given by the equations

$$\frac{\partial g}{\partial t_1} = \cdots = \frac{\partial g}{\partial t_d} = 0.$$

The codimension of V_X inside of $\mathbb{C}^{d+n-1}$ (the neighborhood with coordinates t_i and z_i) is therefore given by the rank of the matrix

$$\begin{bmatrix} \frac{\partial^2 g}{\partial t_1^2} & \frac{\partial^2 g}{\partial t_1 t_2} & \cdots & \frac{\partial^2 g}{\partial t_1 t_d} \\ \frac{\partial^2 g}{\partial t_1 t_2} & \frac{\partial^2 g}{\partial t_2^2} & \cdots & \frac{\partial^2 g}{\partial t_2 t_d} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 g}{\partial t_1 t_d} & \frac{\partial^2 g}{\partial t_2 t_d} & \cdots & \frac{\partial^2 g}{\partial t_d^2} \end{bmatrix}.$$

In our case X is a hypersurface and $V_X \to X$ is an isomorphism, so V_X has dimension d (and recall $d = n-1$). It follows that the above matrix of second order partial derivatives has full rank at all points of V_X.

To finally complete this proof, we return to the map $f\colon Y \to S$. Recall that this is the restriction of $\pi_2\colon W_X \to \widehat{\mathbb{C}P^n}$ to the subspace S. In our choice of coordinates, S is described simply by the vanishing of $z_1, \dots, z_{n-1}$. So the map f looks like

$$t_1, \dots, t_{n-1} \mapsto g(t, 0)$$

in local coordinates about (w, H), and the Hessian at our critical point is just the above matrix of second order partial derivatives. As this matrix has full rank, we are done. $\square$

6. Leftover proofs and geometrical considerations

In this section we tie up some loose ends.

Recall the statement of Proposition 4.6: For any two vanishing cycles δ_i and δ_j, there exists a $g \in G$ such that $g.\delta_i = \pm\delta_j$. We can now give the proof of this, based on our construction of Lefschetz pencils from the last section.

Proof of Proposition 4.6. Let $\widehat{G} = \pi_1(\widehat{\mathbb{C}P^n} \backslash D(X))$, and note that there is a map $G \to \widehat{G}$. There are three geometric facts that go into the proof of this proposition:

(i) The action of G on $H_{n-1}(Y_b)$ factors through an action of $\widehat{G}$;
(ii) The loops w_i and w_j become conjugate in $\widehat{G}$;
(iii) The map of groups $h\colon G \to \widehat{G}$ is surjective.

Accepting these three facts, here is what we do. By (ii) and (iii) together there is a $u \in G$ such that uw_i and $w_j u$ become the same element after applying h. Using (i) this then implies that

$$u.(w_i.x) = w_j.(u.x)$$

for all $x \in H_{n-1}(Y_b)$. Applying the Picard–Lefschetz formula to both sides and simplifying, we get the formula

$$\langle x, \delta_i \rangle \cdot (u.\delta_i) = \langle u.x, \delta_j \rangle \delta_j$$

for all $x \in H_{n-1}(Y_b)$. From here the proof is just algebraic manipulation.

If $\delta_i = 0$ and $\delta_j \neq 0$, then we have $\langle u.x, \delta_j \rangle = 0$ for all $x \in H_{n-1}(Y_b)$. Since u acts on $H_{n-1}(Y_b)$ as an automorphism, this says that δ_j is orthogonal to all of $H_{n-1}(Y_b)$, which is a contradiction. So if $\delta_i = 0$ then $\delta_j = 0$ as well, and in this case there is nothing to prove as we may take $g = id$.

So suppose $\delta_i \neq 0$. Then there exists an $x \in H_{n-1}(Y_b)$ such that $\langle x, \delta_i \rangle \neq 0$. We then get

$$u.\delta_i = \frac{\langle u.x, \delta_j \rangle}{\langle x, \delta_i \rangle} \cdot \delta_j.$$

Abbreviate the coefficient to c, so that $u.\delta_i = c \cdot \delta_j$. Note that $c \neq 0$, as $u.(-)$ is an automorphism. Then

$$c = \frac{\langle u.x, \delta_j \rangle}{\langle x, \delta_i \rangle} = \frac{\langle u.x, \delta_j \rangle}{\langle u.x, u.\delta_i \rangle} = \frac{\langle u.x, \delta_j \rangle}{\langle u.x, c\delta_j \rangle} = \frac{1}{c}.$$

So $c = \pm 1$, and this completes the proof.

It only remains to justify facts (i)–(iii). Fact (i) is the simplest. Recall the space

$$W = \{(x, f) \in \mathbb{C}P^N \times \widehat{\mathbb{C}P^N} \mid f(x) = 0\}.$$

and the map $\pi_2 \colon W \to \widehat{\mathbb{C}P^N}$. Also consider the square of inclusions

$$\begin{tikzcd}
\mathbb{C}P^1 \arrow[r, tail, "j_1"] & \widehat{\mathbb{C}P^N} \\
\mathbb{C}P^1 - \{t_1, \dots, t_k\} \arrow[u, tail, "j_3"] \arrow[r, tail, "j_2"'] & \widehat{\mathbb{C}P^N} \backslash D(X). \arrow[u, tail, "j_4"']
\end{tikzcd}$$

Our Lefschetz pencil came from pulling back $W \to \widehat{\mathbb{C}P^N}$ along j_1, and the monodromy action came about by further pulling back along j_3. But this is the same as pulling back along j_4 and then j_2, and pulling back along j_4 gives a fiber bundle with monodromy action by the group $\pi_1(\widehat{\mathbb{C}P^N} \backslash D(X))$. This justifies the claim.

The proof of conditions (ii) and (iii) will take up the rest of this section. These conditions are restated and proven as Proposition 6.4 below. $\square$

6.1. π_1-considerations. We begin with a lemma that is useful in a variety of circumstances.

Lemma 6.2. *Let X be a smooth, quasi-projective variety over $\mathbb{C}$ and let $Z \hookrightarrow X$ be a closed subvariety that has codimension at least k near every point. Then the inclusion $X \backslash Z \hookrightarrow X$ is a $(2k-1)$-equivalence—i.e., it is an isomorphism on $\pi_i(-)$ for $i < 2k - 1$ and a surjection on $\pi_{2k-1}(-)$.*

Sketch of proof. First assume that Z is smooth. Then by a choosing a tubular neighborhood U of Z, we have a homotopy pushout digram

$$\begin{tikzcd} U\backslash Z \arrow[r, tail] \arrow[d, tail] & X\backslash Z \arrow[d] \\ U \arrow[r, tail] & X. \end{tikzcd}$$

The map $U\backslash Z \hookrightarrow U$ is homeomorphic to $S(N) \hookrightarrow D(N)$ where $S(N)$ and $D(N)$ are the sphere- and disk- bundles of the normal bundle to Z in X. The projection $S(N) \to Z$ has fiber $\mathbb{C}^k - 0 \simeq S^{2k-1}$, and so $S(N) \to Z$ is a $(2k-1)$-equivalence. Then the same is true for $S(N) \to D(N)$, since $D(N) \to Z$ is a weak equivalence. By the Blakers-Massey theorem, $X\backslash Z \hookrightarrow X$ is therefore also a $(2k-1)$-equivalence.

Now consider the general case where Z is not necessarily smooth. Then Z has a filtration

$$Z = Z_0 \supseteq Z_1 \supseteq Z_2 \supseteq \cdots \supseteq Z_n \supseteq Z_{n+1} = \emptyset$$

in which each $Z_i\backslash Z_{i+1}$ is smooth and Z_{i+1} has codimension at least 1 in Z_i near all points.

Consider the associated filtration

$$X\backslash Z = X\backslash Z_0 \subseteq X\backslash Z_1 \subseteq \cdots \subseteq X\backslash Z_n \subseteq X. \tag{6.3}$$

Since $Z_n \hookrightarrow X$ has codimension at least $k+n$ near each point, $X\backslash Z_n \hookrightarrow X$ is a $(2(k+n)-1)$-equivalence. Likewise, $Z_{n-1}\backslash Z_n \hookrightarrow X\backslash Z_n$ is smooth and has codimension at least $k+n-1$ near each point; so the inclusion $X\backslash Z_{n-1} \hookrightarrow X\backslash Z_n$ is a $(2(k+n-1)-1)$-equivalence. Continuing in this way, we find that each link in the chain of inclusions (6.3) is a $(2k-1)$-equivalence, and therefore so is the composite. □

Let $Z \hookrightarrow \mathbb{C}P^N$ be a hypersurface, possibly singular. Recall that a projective line $L \subseteq \mathbb{C}P^N$ meets Z *in general position* if L does not intersect the singular set of Z and if L meets Z transversallly at all points of intersection. Note that topologically $L \cong S^2$, and the complex structure on L equips it with a natural orientation.

Let $b \in \mathbb{C}P^N\backslash Z$. By an "elementary loop" in $L\backslash Z$ we mean a loop based at b that moves straight towards an intersection point $p \in L \cap Z$, runs once around this intersection point counterclockwise (with respect to the orientation of L), and then runs straight back to b. That is, the loop has the form $l^{-1}\sigma l$ where l is a path and σ is a small loop around the point p.

Proposition 6.4. *Suppose that L is a projective line through b that meets Z in general position. Then*

(a) The map $\pi_1(L\backslash Z, b) \to \pi_1(\mathbb{C}P^N\backslash Z, b)$ is surjective.

(b) Let L_1 and L_2 be two lines through b meeting Z in general position (including the possibility that $L_1 = L_2$). If v_1 and v_2 are elementary loops in $L_1\backslash Z$ and $L_2\backslash Z$, then v_1 and v_2 are conjugate in $\pi_1(\mathbb{C}P^N\backslash Z)$.

Proof. The point $b \in \mathbb{C}P^N$ may be regarded as a line ℓ in $\mathbb{C}^{N+1}$. Choose any hyperplane $V \subseteq \mathbb{C}^{N+1}$ which doesn't contain this line. Then points in $\mathbb{P}(V)$ are in bijective correspondence with planes in $\mathbb{C}^{N+1}$ containing ℓ, or equivalently with projective lines in $\mathbb{C}P^N$ containing b. For brevity write $\mathcal{P} = \mathbb{P}(V)$; this is the parameter space for lines through b. For $w \in \mathcal{P}$ write L_w for the corresponding projective line in $\mathbb{C}P^N$.

Let $C \subseteq \mathcal{P}$ be the subspace of lines which are *not* in general position with respect to Z. This is a closed algebraic subvariety, and it is not equal to the entire space $\mathcal{P}$. Since $\mathcal{P}$ is irreducible the codimension of C is at least one near every point. Write $\mathcal{P}_{gp} = \mathcal{P}\backslash C$; this is the open subvariety of lines through b that are in general position with respect to Z.

Define

$$\begin{aligned} E &= \{(x,w) \in \mathbb{C}P^N \times \mathcal{P}_{gp} \,|\, x \in L_w \text{ and } x \notin Z\}, \\ E' &= \{(x,w) \in \mathbb{C}P^N \times \mathcal{P} \,|\, x \in L_w \text{ and } x \notin Z\}. \end{aligned}$$

Then $E \subseteq E'$ is an open subvariety, and $\pi_2 \colon E \to \mathcal{P}_{gp}$ is a fiber bundle with fiber $L\backslash Z$. Note that $E = E'\backslash \pi_2^{-1}(C)$, and so by Lemma 6.2 the inclusion $E \hookrightarrow E'$ is a 1-equivalence. In particular, it is surjective on π_1.

Let us further explain the significance of the spaces E and E'. Denote the projection $E' \to \mathbb{C}P^N\backslash Z$ by p. This is not a fiber bundle, but it is a simple map to understand. For every $x \in \mathbb{C}P^N\backslash(Z \cup \{b\})$ the fiber $p^{-1}(x)$ consists of exactly one point, whereas $p^{-1}(b) \cong \mathcal{P}$. The space E' is the blow-up of $\mathbb{C}P^N\backslash Z$ at the point b.

Consider the diagram below:

$$\begin{tikzcd} \pi_1(L\backslash Z) \arrow[r] \arrow[ddr] & \pi_1(E) \arrow[r, "(\pi_2)_*"] \arrow[d, "j_*"] & \pi_1(\mathcal{P}_{gp}) \\ & \pi_1(E') \arrow[d, "p_*"] & \\ & \pi_1(\mathbb{C}P^N\backslash Z). & \end{tikzcd}$$

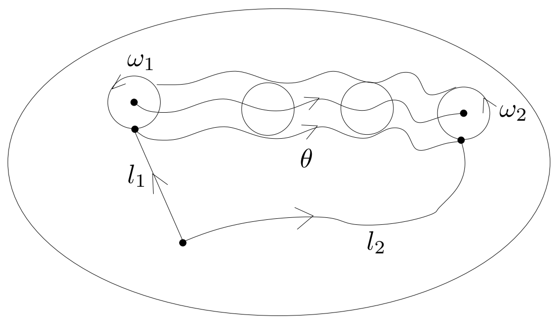
The horizontal row is part of the homotopy long exact sequence for the fiber bundle $L\backslash Z \to E \to \mathcal{P}_{gp}$. Note that $\pi_2 \colon E \to \mathcal{P}_{gp}$ has a splitting χ given by $\chi(w) = (b, w)$ (because all the lines in $\mathcal{P}_{gp}$ automatically contain b by definition). It follows at once that $(\pi_2)_*$ is a split surjection.

We have already seen that j_* is surjective, and we claim the same is true for p_*. To see this it is convenient to temporarily use a basepoint in $\mathbb{C}P^N\backslash Z$ different from b, say b'. Every loop in $\mathbb{C}P^N\backslash Z$ based at b' can be deformed so that it avoids b. But $E'\backslash p^{-1}(b) \to \mathbb{C}P^N\backslash(Z \cup \{b\})$ is an isomorphism. So any loop in $\mathbb{C}P^N\backslash Z$ that avoids b has a unique lifting to E'. This proves that p_* is surjective when the basepoint is b', and of course it must be the same when the basepoint is b. [Alternatively, use the van Kampen theorem together with the fact that $\mathbb{C}P^N\backslash Z$ is the homotopy pushout of $* \leftarrow \mathcal{P} \to E'$.]

It is now a diagram chase to see that the diagonal map is surjective. Let $\alpha \in \pi_1(\mathbb{C}P^N\backslash Z)$ and lift this to an element $\beta \in \pi_1(E)$. Then $\beta \cdot [\chi_*(\pi_2)_*(\beta)]^{-1}$ is in the kernel of $(\pi_2)_*$, so it is the image of an element $\gamma \in \pi_1(L\backslash Z)$. Because $p \circ \chi$ is the trivial map, one readily checks that γ maps to α. This completes the proof of (a).

The proof of (b) is simpler. Let A consist of all points $x \in Z - \{b\}$ with the property that the projective line containing b and x is in general position with respect to Z. Then $A \subset Z$ is an algebraic subvariety. Since Z is irreducible, A has codimension at least one near all of its points. By Lemma 6.2 one sees that $Z\backslash A$ is path-connected.

Recall that we start with two elementary loops v_i in $L_i\backslash Z$, for $i = 1, 2$. Let c_i be the point in $L_i \cap Z$ that v_i encircles. Since $c_1, c_2 \in Z\backslash A$, there is a path σ in $Z\backslash A$ from c_1 to c_2. For each t, the projective line joining b to $\sigma(t)$ intersects Z in a discrete set of points, one of which is $\sigma(t)$ itself. With a little thought one sees that it is possible to choose a map $J\colon D^2 \times I \to \mathbb{C}P^N$ such that on $\{0\} \times I$ this is just σ and such that $(D^2 - \{0\}) \times I$ maps into $\mathbb{C}P^N\backslash Z$. After shrinking, rotating, or deforming things appropriately one gets something like the following picture, where our elementary loops are $v_1 = l_1^{-1}\omega_1 l_1$ and $v_2 = l_2^{-1}\omega_2 l_2$:



Let θ be the restriction of J to $\{p\} \times I$ for some appropriate point $p \in \partial D^2$. Then ω_1 is homotopic to $\theta^{-1}\omega_2\theta$, and therefore

$$v_1 = l_1^{-1}\omega_1 l_1 \simeq l_1^{-1}(\theta^{-1}\omega_2\theta)l_1 \simeq (l_1^{-1}\theta^{-1}l_2)(l_2^{-1}\omega_2 l_2)(l_2^{-1}\theta l_1) = hv_2h^{-1}$$

where $h = l_1^{-1}\theta^{-1}l_2$. $\square$

We have now completed the proof of Proposition 4.6.

7. Proof of the variation formula

Our final task is to prove Proposition 2.8. Recall that this is a statement about the variation inside the bounded sum-of-squares mapping, and that it was the key step in the analysis of the monodromy of Lefschetz pencils.

Let us review the setting. We fix $\rho > 0$ and $\epsilon > 0$ such that $\rho < \epsilon^2$, and define

$$E = \Big\{(z_1, \dots, z_n) \in \mathbb{C}^n \;\Big|\; \sum_j |z_j|^2 \le \epsilon^2 \text{ and } |z_1^2 + \cdots + z_n^2| \le \rho\Big\}$$

and

$$B = \big\{w \in \mathbb{C} \;\big|\; |w| \le \rho\big\}.$$

Define the map $f\colon E \to B$ to send $(z_1, \dots, z_n)$ to $z_1^2 + \cdots + z_n^2$. This is a fiber bundle when restricted to $B - 0$, with the fiber a manifold-with-boundary that is isomorphic to $(D(TS^{n-1}), S(TS^{n-1}))$ (the disk bundle and sphere bundle of the tangent bundle of S^{n-1}). Let $\gamma\colon I \to B - 0$ be the path $\gamma(t) = e^{2\pi i t}$. There is an associated relative variation map

$$\mathrm{Var}_\gamma^{rel}\colon H_{n-1}(F, \partial F) \to H_n(E, F),$$

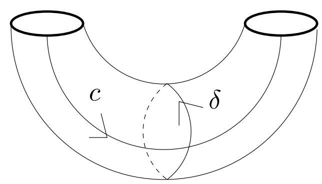
and our goal is to calculate this. Both the domain and codomain are isomorphic to $\mathbb{Z}$.

Let $(\Delta, \delta)\colon (D^n, S^{n-1}) \to (E, F)$ consists of a vanishing cycle δ and corresponding thimble Δ. Proposition 2.8 consists of the formula

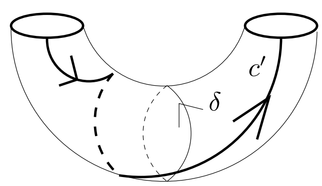
$$\mathrm{Var}_\gamma(x) = (-1)^{\binom{n}{2}} \langle x, \delta\rangle \cdot \Delta$$

where x is any element of $H_{n-1}(F, \partial F)$. Because Var_γ is a map $\mathbb{Z} \to \mathbb{Z}$, it suffices for us to verify the formula on any chosen nonzero element.

The verification of this formula will be easier to understand if we first explore the case $n = 2$. We will freely use the homeomorphism of pairs $(F_\rho, \partial F_\rho) \cong (D(TS^{n-1}), S(TS^{n-1}))$. From now on we'll just write $(F, \partial F)$ for $(F_\rho, \partial F_\rho)$. When $n = 2$ we have $\partial F \cong V_2(\mathbb{R}^2) = O(2) \cong S^1 \amalg S^1$, and $F \cong S^1 \times I$ (because the tangent bundle to S^1 is trivial). We have the following picture, which shows a vanishing cycle $\delta \in H_1(F)$ and also a cycle $c \in H_1(F, \partial F)$ such that $\langle c, \delta\rangle = 1$:



We claim that as this picture moves through the monodromy it becomes the following:



So the cycle c changes to $c' = c + \delta$.

???

Let us work through the definition of $\mathrm{Var}^{rel}_\gamma$ in this case. We construct the lifting λ by

$$\lambda(z_1, \ldots, z_n, t) = (e^{\pi i t} z_1, \ldots, e^{\pi i t} z_n).$$

Consider the diagram

$$\begin{tikzcd}
 & H_1(F, \partial F) \arrow[r] & H_2(F \times I, \partial(F \times I)) \arrow[r, "\lambda_*"] \arrow[dd, "\partial"] & H_2(E, F \cup E') \arrow[dd, "\cong"', "\partial"] & H_2(E, F) \arrow[l, "\cong"'] \arrow[dd, "\cong"', "\partial"] \\
H_1(D^1, \partial D^1) \arrow[r] \arrow[ur, "C_*"] & H_2(S, \partial S) \arrow[ur, "(C \times id)_*"'] \arrow[dd] & & & \\
 & & H_1(\partial(F \times I)) \arrow[r, "\lambda_*"] & H_1(F \cup E') & H_1(F) \arrow[l, "\cong"'] \\
 & H_1(\partial S) \arrow[ur] & & &
\end{tikzcd}$$

where $S = D^1 \times I$ and we are regarding $C\colon (D^1, \partial D^1) \to (F, \partial F)$ as our model for c (i.e., c is the pushforward under C_* of the fundamental class in $H_1(D^1, \partial D^1)$). Then $\mathrm{Var}^{rel}_\gamma(c)$ is the image of c across the top row, and we can compute this by instead computing its boundary. The class c itself is the image of the canonical element in $H_1(D^1, \partial D^1)$, and so we are reduced to understanding the map $\partial S = \partial(D^1 \times I) \to F \cup E'$ obtained by restricting λ. On $D^1 \times \{0\}$ this is the original map $D^1 \to F$ representing c. On $D^1 \times \{1\}$ it is the composite of c with the monodromy map $F \to F$. On $(\partial D^1) \times I$ the map is simply $(a, t) \mapsto e^{\pi i t} \cdot c(a)$.

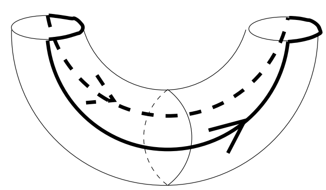
To explain this a bit better we now change from $(F, \partial F)$ to the model $(D(TS^{n-1}), S(TS^{n-1}))$. Let $f \colon \partial(D^1 \times I) \to F \cup E'$ be the map we are studying. In this new model the vanishing cycle is the 0-section $S^{n-1} \to D(TS^{n-1})$ given by $u \mapsto (u, 0)$. We can model c by a path which starts with a 2-frame (x, y) and then slowly changes y to $-y$ by the evident straight line passing through 0. This path is $f|_{D^1 \times \{0\}}$.

The monodromy action $F \to F$ becomes the map $D(TS^{n-1}) \to D(TS^{n-1})$ given by $(u, v) \mapsto (-u, -v)$. So $f|_{D^1 \times \{1\}}$ is the path that starts with the 2-frame $(-x, -y)$ and slowly moves $-y$ to y along the straight-line path. Note that $V_2(\mathbb{R}^2)$ consists of two circles, corresponding to the positively- and negatively-oriented frames. The frame $(-x, -y)$ is obtained by rotating (x, y) through 180 degrees, and in terms of our picture this will be on the same circle as (x, y) but on the opposite side.

The analysis of $f_0 = f|_{\{0\} \times I}$ and $f_1 = f|_{\{1\} \times I}$ is slightly more confusing. In terms of the $(F, \partial F)$ model (rather than the disk bundle/sphere bundle model), the former is a path that starts at some point $(z_1, \ldots, z_n) = x + iy$ with $x \cdot y = 0$ and $|x|^2 = \rho + |y|^2$, and then progresses as $t \mapsto (e^{\pi i t} z_1, \ldots, e^{\pi i t} z_n)$. The path f_1 is a similar path that starts at $(\bar{z}_1, \ldots, \bar{z}_n) = x - iy$ and does the same thing. We would like to translate this into our disk bundle/sphere bundle model, but there is a difficulty in that this model is really only valid on the fiber over the basepoint—whereas our paths f_0 and f_1 are mostly not in this fiber. This is the difference between the map we have, of the form $\partial S \to F \cup E'$, and the map we would like which would have the form $\partial S \to F$.

????

We get the following picture:



Now we tackle the general case:

Proof of Proposition 2.8. Fix a representative

$$\begin{array}{ccc} S^{n-2} & \xrightarrow{c'} & \partial F \\ \downarrow & & \downarrow \\ D^{n-1} & \xrightarrow{c} & F \end{array}$$

for the generator of $H_{n-1}(F, \partial F)$. We start again with the diagram

$$\begin{tikzcd}
& H_{n-1}(F,\partial F) \arrow[r] & H_n(F\times I,\partial(F\times I)) \arrow[r, "\lambda_*"] \arrow[dd, "\partial"] & H_n(E, F\cup E') \arrow[dd, "\partial", "\cong"'] & H_n(E,F) \arrow[l, "\cong"'] \arrow[dd, "\partial", "\cong"'] \\
H_{n-1}(D,\partial D) \arrow[ur, "{(c,c')}"] \arrow[r] & H_n(S,\partial S) \arrow[ur] \arrow[dd] & & & \\
& & H_{n-1}(\partial(F\times I)) \arrow[r, "\lambda_*"] & H_{n-1}(F\cup E') & H_{n-1}(F) \arrow[l, "\cong"'] \\
& H_{n-1}(\partial S) \arrow[ur] & & &
\end{tikzcd}$$

where $D = D^{n-1}$ and $S = D \times I$. Let $R\colon E' \to F$ denote the retraction, and let $r\colon F \to S^{n-1}$ denote the map ?????. Finally, let h denote the composite map

$$\partial S = \partial(D^{n-1} \times I) \to F \cup E' \xrightarrow{R} F \xrightarrow{r} S^{n-1}.$$

The result will follow after we show two things:

(1) Equipping D^{n-1} with its standard orientation, and taking the induced orientation on ∂S, the degree of the map h is -1.
(2) The map $H_{n-1}(D, \partial D) \to H_{n-1}(F, \partial F)$ sends the canonical generator to an element c having the property that $\langle c, \delta\rangle = (-1)^{\binom{n}{2}}$.

What makes this process manageable is the ability to write down an explicit formula for R. We will give a map $R'\colon E' \to V_2(\mathbb{R}^n)$, and then the map R will be $q \circ R'$ where q is our standard homeomorphism between $V_2(\mathbb{R}^n)$ and F. To describe the formula for R', let $z \in \mathbb{C}^n$ and write $z = x + iy$ for $x, y \in \mathbb{R}^n$. There exists an $\alpha \in \mathbb{R}$ such that the real and imaginary parts of $e^{-i\alpha}z$ are orthogonal, and α is well-defined up to integral multiples of π. (In fact, if $\sum z_j^2 = re^{i\theta}$ then $\alpha \equiv \theta/2 \mod \pi$.) If we write $e^{-i\alpha}z = x' + iy'$, define $R'(z)$ to be the 2-frame obtained by taking the real and imaginary parts of

$$e^{i\alpha}\left[\frac{x'}{|x'|} + i\frac{y'}{|y'|}\right].$$

Note that altering α by an odd multiple of π changes the signs on x' and y' AND changes the sign on $e^{i\alpha}$, and therefore has no effect on $R'(z)$. The main property we need about R' is the following:

> If $x, y \in \mathbb{R}^n$ are orthogonal and $z = e^{it}(x + iy)$, then $R'(z)$ is obtained by rotating the 2-frame $(\frac{x}{|x|}, \frac{y}{|y|})$ counterclockwise through t radians..

The map $S^{n-2} \to \partial F$ may be modelled by $v \mapsto (e_1, v)$, where e_1 is the first standard basis element of $\mathbb{R}^n$ and S^{n-2} is regarded as the sphere perpendicular to e_1. The map $D^{n-1} \to F$ is then described by

$$tv \mapsto (e_1, tv)$$

where $v \in S^{n-2}$ and $t \in I$. The retraction $r\colon F \to S^{n-1}$ is the map $(u, v) \to u$.

We now proceed to analyze the

$$\partial(D^{n-1} \times I) \to F \cup E' \to F \to S^{n-1}.$$

On $D^{n-1} \times \{0\}$ this is the map $(tv, 0) \to (e_1, tv) \to e_1$. On $D^{n-1} \times \{1\}$ this is $(tv, 1) \to (-e_1, -tv) \to -e_1$. Finally, on $(\partial D^{n-1}) \times I$ it is the map

$$(v, t) \longrightarrow e^{\pi i t}(e_1 + iv) \xrightarrow{r \circ R'} \cos(\pi t)e_1 - \sin(\pi t)v.$$

It follows that $h^{-1}(-e_2)$ is a singleton set consisting of the point $(e_2, \frac{1}{2})$. We need to compute the local degree of h near this point. If we write $v \in S^{n-2}$ as $v = (v_2, v_3, \ldots, v_n)$, then $v_3, \ldots, v_n, t$ give a positively-oriented coordinate system for $\partial(D^{n-1} \times I)$ near the point $(e_2, \frac{1}{2})$. Likewise $u_1, u_3, \ldots, u_n$ give a positively-oriented coordinate system on the target S^{n-1} around the point $-e_2$. In these coordinates one has that

$$h(v_3, \ldots, v_n, t) = (\cos(\pi t), -\sin(\pi t)v_3, -\sin(\pi t)v_4, \ldots, -\sin(\pi t)v_n).$$

The Jacobian is the $(n-1) \times (n-1)$ matrix

$$J = \begin{bmatrix} 0 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\pi & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

whose determinant is $-\pi$. This shows that the degree of h is -1.

To prove point (2), recall that the vanishing cycle δ can be represented by the map $S^{n-1} \to F$ given by $u \mapsto (u, 0)$ (it is the 0-section of the tangent bundle for S^{n-1}). The relative cycle (c, c') is represented by $D^{n-1} \to F$ given by $tv \mapsto (e_1, tv)$. Clearly the only intersection between the images of these two maps is when $u = e_1$ and $v = 0$. To determine $\langle c, \delta \rangle$ we therefore only need to compute an intersection multiplicity, by juxtaposing oriented bases for the tangent space to c and the tangent space to δ and comparing the resulting basis to the chosen orientation of F.

The disk bundle model for F consists of points (u, v) such that $|u| = 1$, $|v| \leq 1$, and $u \cdot v = 0$. Near the point $(e_1, 0)$ we choose local coordinates $u_2, \ldots, u_n, v_2, \ldots, v_n$. The orientation for F comes from the complex structure, however, and that tells us that $u_2, v_2, u_3, v_3, \ldots, u_n, v_n$ would be an oriented coordinate system for F.

The tangent space to δ at $(e_1, 0)$ has coordinates $u_2, \ldots, u_n$. The tangent space to (c, c') at $(e_1, 0)$ has coordinates $v_2, \ldots, v_n$. So to compute the intersection multiplicity we compare the two coordinate systems

$$[v_2, v_3, \ldots, v_n, u_2, u_3, \ldots, u_n] \quad \text{and} \quad [u_2, v_2, \ldots, u_n, v_n]$$

The number of transpositions needed to move from the first to the second is

$$(n-1) + (n-2) + (n-3) + \cdots + 1 = \tfrac{n(n-1)}{2}.$$

So $\langle c, \delta \rangle = (-1)^{\binom{n}{2}}$. $\square$

CHAPTER 7

Deligne's proof of the Riemann hypothesis

In this chapter we present Deligne's first proof.

1. Grothendieck L-functions

L-functions are generalizations of zeta functions. Like zeta functions, one starts with an algebraic or number-theoretic object and from this data constructs—in some way—an analytic function defined on a portion of the complex plane. Typically this function will decompose as an infinite product, have an analytic continuation to the entire complex plane, and certain "special values" of this analytic continutation will encode interesting information about the original object. There is no abstract definition of the class of mathematical objects called "L-functions". Rather, there is a collection of examples which arise in different contexts, and in some cases deep conjectures about how some examples relate to others.

In the context of algebraic geometry over finite fields, Grothendieck's L-functions are a very mild generalization of the zeta functions we have already seen—essentially it is just the generalization from constant coefficents to twisted coefficients that one is familiar with from topology. These L-functions play a crucial role in Deligne's proof, however. His proof works by reducing the Riemann hypothesis for zeta functions of varieties to a related claim about more general L-functions over *curves*.

Strictly speaking, we could probably give the relevant facts about Grothendieck's L-functions fairly quickly and be done with it. But in order to set this material into a larger context, we will first give a brief—in fact, very brief—overview of the L-functions from analytic number theory that serve as their prototypes.

1.1. The Riemann zeta function. Recall one has $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ and this also admits a product description as

$$\zeta(s) = \prod_{p} (1 - p^{-s})^{-1}$$

where p ranges over all primes in $\mathbb{Z}$. The factors $(1 - p^{-s})^{-1}$ are called **Euler factors**.

1.2. Dirichlet L-functions. Fix a positive integer $m > 1$ and a homomorphism $\chi\colon (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{C}^*$ (called a *character* of the group $(\mathbb{Z}/m\mathbb{Z})^*$). Extend χ to a function $\tilde{\chi}\colon \mathbb{Z} \to \mathbb{C}$ by

$$\tilde{\chi}(k) = \begin{cases} 0 & \text{if } (k,m) \neq 1, \\ \chi([k]) & \text{otherwise, where } [k] \text{ is the reduction of } k \text{ mod } m. \end{cases}$$

One can check that $\chi(kl) = \chi(k)\chi(l)$ for all integers k and l.

Define the Dirichlet L-function as $L_\chi(s) = \sum_{n=1}^{\infty} \chi(n)n^{-s}$ and check that one can also write this as

$$L_\chi(s) = \prod_p (1 - \chi(p)p^{-s})^{-1}.$$

One proves that this function is analytic in the range $Re(s) > 0$, and admits an analytic continuation to the entire complex plane.

1.3. Hecke L-functions. ?????

1.4. Artin L-functions. Here one starts with a Galois extension L/K of number fields. Write $\mathcal{O}_K \hookrightarrow \mathcal{O}_L$ for the rings of integers in K and L. Given a prime ideal p in $\mathcal{O}_K$, its expansion into $\mathcal{O}_L$ will typically not be prime: instead it will factor (uniquely) as

$$p = q_1^{e_1} \cdots q_g^{e_g}$$

where each q_i is a prime in $\mathcal{O}_L$. Here is a list of facts one can prove about this decomposition:

(1) Let $n = [L : K] = \#\mathrm{Gal}(L/K)$. Applying $N_{L/K}$ to the above equation gives
$$p^n = N_{L/K}(q_1)^{e_1} \cdots N_{L/K}(q_g)^{e_g}.$$
By uniqueness of prime factorization of ideals, for each i one must have $N_{L/K}(q_i) = p^{f_i}$ for some $f_i \geq 1$.

(2) One can prove that f_i is the degree of the field extension $[\mathcal{O}_L/q_i : \mathcal{O}_K/p]$.

(3) $\mathrm{Gal}(L/K)$ acts transitively on the q_i's, therefore $e_1 = e_2 = \cdots = e_g$ and $f_1 = f_2 = \cdots = f_g$. So let us drop the subscripts and just write e and f. The number e is called the **ramification degree** of p, and f is called the **residue field degree** of p. When $e = 1$ the prime p is called **unramified**.

(4) Substituting $e_i = e$ and $f_i = f$ into the equation equation from (1) we find that $efg = n$.

(5) For each prime q of $\mathcal{O}_L$, let $G_q \hookrightarrow G$ be the subgroup of automorphisms which fix q. This is called the **decomposition subgroup** for the prime q. Since G acts transitively on the q_i's (a set of size g), we have that $\#G_q = \frac{\#G}{g} = \frac{n}{g} = ef$.

(6) There is a homomorphism $G_q \to \mathrm{Gal}(L_q/K_p)$, and one can prove that it is surjective. The kernel is called the **inertia group** of q, and written I_q:
$$1 \to I_q \to G_q \to \mathrm{Gal}(L_q/K_p) \to 1.$$

(7) $\#I_q = \frac{\#G_q}{\#\mathrm{Gal}(L_q/K_p)} = \frac{ef}{f} = e$.

(8) L_q/K_p is an extension of finite fields, so the group $\mathrm{Gal}(L_q/K_p)$ is cyclic and is generated by the Frobenius homomorphism $x \mapsto x^{(Np)^f}$ (where $Np = \#K_p$). A preimage of this element in G_q is called a **Frobenius element** for q over p and denoted $\phi_{q/p}$. Note that if p is unramified then $\#I_q = e = 1$ and therefore this Frobenius element is uniquely defined.

(9) If $\sigma \in G$ and $\sigma(q) = q'$ then $\phi_{q'/p} = \sigma\phi_{q/p}\sigma^{-1}$. (In particular, if G is abelian then the Frobenius element $\phi_{q/p}$ does not depend on the choice of q.)

Now let V be a representation of G acting on a complex vector space. Let $\rho\colon G \to \mathrm{Aut}(V)$ denote the action map. Define a function $L_V(s)$ by

$$L_V(s) = \prod_{p \subseteq \mathcal{O}_K} \det(id - (Np)^{-s} \cdot \rho(\phi_{q/p}))^{-1}$$

where the product is over all unramified primes of $\mathcal{O}_K$ and for each such prime one chooses a prime $q \subseteq \mathcal{O}_L$ lying over it. Note that each of the local factors $\det(id - (Np)^{-1}\rho(\phi_{q/p}))^{-1}$ seems to depend on the choice of prime q, but by fact (9) above it does not: different choices of q's give rise to conjugate Frobenius elements, which will have the same characteristic polynomial.

The function $L_V(s)$ is not quite the Artin L-function attached to the representation V, but it is close. To construct the Artin L-function one has to add to the above product certain factors for each of the ramified primes of $\mathcal{O}_K$, and the description of these factors is more complicated—it would be too much of a distraction at the moment. The above definition is good enough to get the general idea.

Observe that the notion of Artin L-function generalizes that of Dirichlet L-function. If $L = \mathbb{Q}(\mu_m)$, then $\mathrm{Gal}(L/\mathbb{Q}) \cong (\mathbb{Z}/m\mathbb{Z})^*$. A representation of this Galois group on $\mathbb{C}$ is precisely given by a character, and the Dirichlet L-function for this character corresponds with the Artin L-function (is this true?)

1.5. Grothendieck L-functions. We now come to the case we really care about. Let X be a scheme over a finite field $\mathbb{F}_q$. Let $\mathcal{F}$ be a constructible $\mathbb{Q}_l$-sheaf on X. Let $F\colon X \to X$ be the Frobenius morphism.

???

Now one defines

$$L_{X,\mathcal{F}}(t) = \prod_{x\in|X|} \det(id - t^{\deg(x)}F_x^*|\mathcal{F}_x)^{-1}.$$

Two observations are worth making right away. First, if $0 \to \mathcal{F}' \to \mathcal{F} \to \mathcal{F}'' \to 0$ is a short exact sequence of sheaves then one has

$$L_{X,\mathcal{F}}(t) = L_{X,\mathcal{F}'}(t) \cdot L_{X,\mathcal{F}''}(t).$$

Second, if $\mathcal{F}$ is the constant sheaf $\mathbb{Q}_l$ then the local factors just reduce to $(1 - t^{\deg(x)})^{-1}$ and we have

$$L_{X,\mathbb{Q}_l}(t) = \prod_{x\in|X|} (1 - t^{\deg(x)})^{-1} = \zeta_X(t).$$

Grothendieck proved a Lefschetz trace formula for L-functions:

Theorem 1.6 (Local trace formula). *For each $n \geq 1$,*

$$\sum_{x\in X(\mathbb{F}_{q^n})} \mathrm{Tr}((F^n)^*;\mathcal{F}_x) = \sum_{i=0}^{\infty} (-1)^i \,\mathrm{Tr}((F^*)^n|_{H_c^i(\overline{X};\mathcal{F})}).$$

Corollary 1.7. *One has the formula*

$$L_{X,\mathcal{F}}(t) = \prod_i \Big[\det(id - t^f F^*|_{H_c^i(\overline{X};\mathcal{F})})\Big]^{(-1)^{i+1}}.$$

2. First reductions of the proof

Fix a smooth, projective variety $X/\mathbb{F}_q$. We will say that "the Riemann hypothesis holds for X in dimension i" if all the eigenvalues of Frobenius on $H^i(\overline{X};\mathbb{Q}_l)$ have absolute norm $q^{i/2}$.

PROPOSITION 2.1. *Fix a finite field $\mathbb{F}_q$ and a prime l such that $(l,q) = 1$. Assume there is a real number $C > 0$ such that for all smooth, even-dimensional, projective varieties X over $\mathbb{F}_q$, the eigenvalues of Frobenius on $H^X(\overline{X};\mathbb{Q}_l)$ are less than $C \cdot q^{X/2}$. Then the Riemann hypothesis holds for all smooth, projective varieties Y over $\mathbb{F}_q$ and all $i \geq 0$.*

The above proposition is telling us three things:

(1) We can restrict to studying the Riemann hypothesis on the middle-dimensional cohomology groups;
(2) We can restrict to looking at even-dimensional varieties;
(3) We do not have to prove the Riemann hypothesis on the nose: it is enough to bound the norms of the eigenvalues by some fixed multiple of $q^{X/2}$.

The proof that these restrictions are sufficient is very easy: it only uses the Künneth and Weak Lefschetz theorems.

PROOF. We first show that for all smooth, projective varieties $X/\mathbb{F}_q$ of dimension d, the Riemann hypothesis holds in dimension d. Let α be an eigenvalue of Frobenius on $H^d(\overline{X};\mathbb{Q}_l)$. Then for every $k > 0$, α^{2k} is an eigenvalue of Frobenius on $H^{2kd}(\overline{X^{2k}};\mathbb{Q}_l)$ by the Künneth Theorem. By hypothesis we therefore have

$$|\alpha|^{2k} = |\alpha^{2k}| \leq C \cdot q^{kd}$$

(and likewise for all the conjugates of α). Taking roots then gives

$$|\alpha| \leq C^{1/2k} \cdot q^{d/2},$$

and taking the limit as $k \mapsto \infty$ gives that $|\alpha| \leq q^{d/2}$.

By Poincaré Duality we know that q^d/α is another eigenvalue of Frobenius on $H^d(\overline{X};\mathbb{Q}_l)$. Applying the above arguments to *this* eigenvalue shows $|q^d/\alpha| \leq q^{d/2}$, or $q^{d/2} \leq |\alpha|$. So in fact we have $|\alpha| = q^{d/2}$, as desired. This completes the first step of the proof.

We have now shown that the Riemann hypothesis holds for all middle-dimensional cohomology groups. We now prove that it holds for all cohomology groups $H^i(\overline{Y};\mathbb{Q}_l)$, by an induction on $\dim Y - i$. The base was just established, and the induction step is taken care of by the Weak Lefschetz Theorem. To be precise, let Y be a smooth, projective variety overy $\mathbb{F}_q$ of dimension d and let $i < \dim Y$. Let $Z \hookrightarrow Y$ be a smooth hyperplane section of Y. By the Weak Lefschetz Theorem, $H^i(\overline{Y};\mathbb{Q}_l) \to H^i(\overline{Z};\mathbb{Q}_l)$ is an isomorphism. But by induction we know the Riemann hypothesis for the eigenvalues of Frobenius on $H^i(\overline{Z};\mathbb{Q}_l)$, and so we are done. $\square$

For the second reduction in the proof we make use of Lefschetz pencils. Let $X/\mathbb{F}_q$ be a smooth, projective variety of dimension d. Assume by induction that we have a bound for the norms of the eigenvalues of Frobenius on middle-dimensional cohomology groups for $(d-1)$-dimensional varieties. We know that there exists a Lefschetz Pencil $f\colon X' \to \mathbb{P}^1$, with X' constructed as a blow-up of X along a $(d-2)$-dimensional subvariety $Z \hookrightarrow X$.

????

There is a spectral sequence

$$E_2^{p,q} = H^p(\overline{\mathbb{P}^1}; R^q f_*(\mathbb{Q}_l)) \Rightarrow H^{p+q}(\overline{X'};\mathbb{Q}_l).$$

We know that $H^p(\overline{\mathbb{P}^1};\mathcal{F}) = 0$ for $p > 2$, so this spectral sequence consists of three vertical lines and a single column of differentials (from the $p = 0$ line to the $p = 2$

line). The Frobenius acts on the entire spectral sequence (commuting with the differentials), and so to analyze the eigenvalues of Frobenius on $H^d(\overline{X'};\mathbb{Q}_l)$ it will be enough to analyze the eigenvalues on the three groups

$$H^0(\overline{P^1};R^df_*(\mathbb{Q}_l)),\quad H^1(\overline{P^1};R^{d-1}f_*(\mathbb{Q}_l)),\quad \text{and}\quad H^2(\overline{P^1};R^{d-2}f_*(\mathbb{Q}_l)).$$

More specifically, the spectral sequence gives a filtration

$$H^d(\overline{X'};\mathbb{Q}_l)=V_0\supseteq V_1\supseteq V_2\supseteq 0$$

that is preserved by the Frobenius map, together with isomorphisms

$$\begin{aligned}
V_0/V_1 &\cong \ker[d\colon H^0(\overline{\mathbb{P}^1};R^df_*(\mathbb{Q}_l))\to H^2(\overline{\mathbb{P}^1};R^{d-1}f_*(\mathbb{Q}_l))]\\
V_1/V_2 &\cong H^1(\overline{P^1};R^{d-1}f_*(\mathbb{Q}_l)),\\
V_2 &\cong \operatorname{coker}[d\colon H^0(\overline{\mathbb{P}^1};R^{d-1}f_*(\mathbb{Q}_l))\to H^2(\overline{\mathbb{P}^1};R^{d-2}f_*(\mathbb{Q}_l))].
\end{aligned}$$

The reduction is accomplished through several applications of the following lemma:

LEMMA 2.2. *Let E be a field and suppose we have a commutative diagram of finite-dimensional E-vector spaces*

$$\begin{tikzcd}
0 \arrow[r] & V' \arrow[r] \arrow[d, "f'"'] & V \arrow[r] \arrow[d, "f"] & V'' \arrow[r] \arrow[d, "f''"] & 0\\
0 \arrow[r] & V' \arrow[r] & V \arrow[r] & V'' \arrow[r] & 0
\end{tikzcd}$$

where the two rows are the same short exact sequence. Then $P_t(f)=P_t(f')\cdot P_t(f'')$.

PROOF. This is elementary. Pick a basis for V' and extend it to a basis for V. Decompose the matrix for f with respect to this basis into the four evident blocks, and compute a determinant. □

At this point the proof breaks up into three pieces. For brevity, write $R^i = R^if_*(\mathbb{Q}_l)$.

The eigenvalues of Frobenius on $\mathbf{H^2}(\overline{\mathbb{P}^1};\mathbf{R^{d-2}})$.

From our knowledge of Lefschetz pencils, the sheaf R^{d-2} in constant on $\mathbb{P}^1$ with fiber $H^{d-2}(Y;\mathbb{Q}_l)$. It follows that

$$H^2(\overline{\mathbb{P}^1};R^{d-2})\cong H^2(\overline{\mathbb{P}^1};\mathbb{Q}_l)\otimes H^{d-2}(Y;\mathbb{Q}_l).$$

We know that Frobenius acts on $H^2(\overline{\mathbb{P}^1};\mathbb{Q}_l)$ with eigenvalue q, and by induction it acts on $H^{d-2}(Y;\mathbb{Q}_l)$ with eigenvalues having absolute norm $q^{(d-2)/2}$ (use Weak Lefschetz!) So the eigenvalues on $H^2(\overline{\mathbb{P}^1};R^{d-2})$ have the required norm.

The eigenvalues of Frobenius on $\mathbf{H^0}(\overline{\mathbb{P}^1};\mathbf{R^d})$.

This case is similar to the previous one, but slightly harder. We again have that the sheaf R^d is constant on $\mathbb{P}^1$, but now with fiber $H^d(Y;\mathbb{Q}_l)$. So

$$H^0(\overline{\mathbb{P}^1};R^d)\cong H^d(Y;\mathbb{Q}_l).$$

But we know the Riemann hypothesis for $H^d(Y;\mathbb{Q}_l)$ by induction, as $\dim Y = d-1$.

????

The eigenvalues of Frobenius on $\mathbf{H^1}(\overline{\mathbb{P}^1};\mathbf{R^{d-1}})$.

This is the crucial, and difficult, case. Let $j\colon U \hookrightarrow \mathbb{P}^1$ be the inclusion. Then $R^{d-1} \cong j_*\mathcal{E} \oplus \mathcal{A}$, where $\mathcal{A}$ is the constant sheaf with stalks $H^{d-1}(F;\mathbb{Q}_l)$. Then

$$H^1(\overline{\mathbb{P}^1};R^{d-1}) \cong H^1(\overline{\mathbb{P}^1};j_*\mathcal{E}) \oplus H^1(\overline{\mathbb{P}^1};\mathcal{A}) \cong H^1_c(\overline{U};\mathcal{E}) \oplus 0 = H^1_c(\overline{U};\mathcal{E}).$$

At this point the proof breaks up into three subcases. Before analyzing these, note that we have at this point reduced the whole question of the Riemann Hypothesis to analyzing the eigenvalues of Frobenius on certain groups $H^1_c(\overline{U};\mathcal{E})$, where U is an open subscheme of $\mathbb{P}^1$ and $\mathcal{E}$ is a certain locally constant sheaf on U. In what follows, we will make use of some further special properties about $\mathcal{E}$.

???????

3. Preliminaries on the symplectic group

Let V be a vector space over a field E, and assume that V is equipped with a skew-symmetric, non-degenerate bilinear form $\langle -,-\rangle$. Skew-symmetric means $\langle v,w\rangle = -\langle w,v\rangle$ and non-degenerate means that $\langle v,-\rangle$ is the zero functional only when $v=0$. A form having these two properties is called a **symplectic form**.

Define $\mathrm{Sp}(V)$ to be the subgroup of $\mathrm{GL}(V)$ consisting of automorphisms that preserve the form: that is,

$$\mathrm{Sp}(V) = \{h \in \mathrm{GL}(V) \,|\, \langle h(v),h(w)\rangle = \langle v,w\rangle \text{ for all } v,w \in V\}.$$

This is called the **symplectic group** of the pair $(V,\langle -,-\rangle)$.

It turns out that a symplectic form can exist on V only when V is even-dimensional, and that up to isomorphism V has only *one* such form. This is easy to explain. Pick any nonzero element $v_1 \in V$. Since the functional $\langle v_1,-\rangle$ is nonzero, it is surjective: so pick a vector $w_1 \in V$ with $\langle v_1,w_1\rangle = 1$. Let $V' = (E.v_1)^\perp \cap (E.w_1)^\perp$. Each of the two perpendicular complements is a hyperplane in V, and they are not equal because v_1 belongs to the former and not to the latter. So V' is a codimension two subspace ov V. The restriction of $\langle -,-\rangle$ to V' is readily checked to be non-degenerate, so repeat the above process for V'. Continuing inductively, one produces a basis $v_1,\dots,v_n,w_1,\dots,w_n$ for V such that

$$\Big(\langle v_i,w_j\rangle\Big)_{i,j} = \begin{bmatrix} \mathbf{O} & \mathbf{I} \\ -\mathbf{I} & \mathbf{O}\end{bmatrix}.$$

If we let J_{2n} denote the block matrix in the above formula, let us define

$$\mathrm{Sp}(2n,E) = \{A \in M_{2n\times 2n}(E) \,|\, AJA^T = J\}.$$

Then this group is isomorphic to $\mathrm{Sp}(V)$ for any $2n$-dimensional symplectic space V.

Now, if G is any group acting (on the left) on a vector space V, then one can form the vector space of coinvariants (or orbit space)

$$V_G = V/\langle v - gv \,|\, v \in V, g \in G\rangle.$$

One can also look at the space V^G of vector invariant under G, but this turns out to be zero in most of the cases we will be interested in below. Instead one can consider *invariant functions* defined on V. To be precise, let $E[V]$ denote the ring of polynomial functions on V. Then G acts on the right on $E[V]$ by $(\phi\cdot g)(v) = \phi(g.v)$. One can then consider the ring of coinvariants

$$E[V]^G = \{\phi \in E[V] \,|\, \phi \cdot g = \phi\}.$$

The classical problems of invariant theory can be loosely stated as follows:

Problem: For familiar groups G and familiar representations V, compute V_G and $E[V]^G$.

This was the subject of Weyl's classic book ?????.

Now let $G = \mathrm{Sp}(V)$. This acts naturally on V, as well as all the tensor powers $V^{\otimes k}$. Notice that $E[V \otimes V]^G$ has an obvious element, namely the function $\phi(v, w) = \langle v, w\rangle$. Generalizing this, suppose we have a partition $\mathcal{P}$ of $\{1, 2, \dots, 2r\}$ into 2-element subsets $\{i_u, j_u\}$ with $i_u < j_u$, for $1 \le u \le r$. Then we can construct a function $\phi_{\mathcal{P}} \in E[V^{\otimes 2r}]^G$ by

$$\phi_{\mathcal{P}}(v_1 \otimes \cdots \otimes v_{2r}) = \prod_u \langle v_{i_u}, v_{j_u}\rangle.$$

Such functions are called *complete contractions*.

Here is the classical theorem we will need:

THEOREM 3.1. *(a) When k is odd one has $[V^{\otimes k}]_G = 0$.*

(b) When k is even, the ring of invariants $E[V]^G$ is the subring of $E[V]$ generated by the complete contractions.

(c) When k is even there is an isomorphism $[V^{\otimes k}]_G \cong E^{???}$ given by ??????.

4. The fundamental estimate

Let U be an open subset of $\mathbb{A}^1$ over the field $\mathbb{F}_q$. Fix $\beta \in \mathbb{Z}$ and let $\mathcal{F}$ be a $\mathbb{Q}_l$-sheaf on U with the following properties:

(1) $\mathcal{F}$ is equipped with a skew-symmetric bilinear form $\psi\colon \mathcal{F} \otimes \mathcal{F} \to \mathbb{Q}_l(-\beta)$ which is non-degenerate on each fiber $\mathcal{F}_x$;
(2) The image of $\pi_1(\bar{U}, u)$ in $\mathrm{GL}(\mathcal{F}_u)$ is an open subgroup of $\mathrm{Sp}(\mathcal{F}_u, \psi_u)$;
(3) For every $x \in |U|$, the polynomial $\det(id - tF_x|\mathcal{F}_x)$ has rational coefficients.

Under these hypotheses we will prove:

PROPOSITION 4.1.

(a) For each $x \in |U|$, the eigenvalues of F_x on $\mathcal{F}_x$ are algebraic numbers of absolute norm $q_x^{\beta/2}$.

(b) The eigenvalues of F on $H^1_c(\bar{U}; \mathcal{F})$ are algebraic numbers, and the norm of each of their conjugates α satisfies

$$|\alpha| \le q^{\frac{\beta}{2}+1}.$$

The proof proceeds by analyzing two descriptions of the L-function $L_{U,\mathcal{F}}(t)$ and playing them off of each other:

$$L_{U,\mathcal{F}}(t) = \prod_{x \in |U|} \det(id - t^{\deg(x)} F_x|\mathcal{F}_x)^{-1} = \prod_i \det(id - tF^*|_{H^i_c(U;\mathcal{F})})^{(-1)^{i+1}}.$$

Note that the product on the left is a power series in $\mathbb{Q}[[t]]$, whereas the product on the right is a rational function in $\mathbb{Q}_l(t)$. Our method will consist in analyzing the radius of convergence for the given analytic function, from the two different perspectives.

The main 'trick', if that is the right word, is to consider the L-functions not just for $\mathcal{F}$ but also for all the even tensor powers of $\mathcal{F}$. It is only for these even tensor powers that gets a close connection between the radius of convergence of the product and that of the local factors.

Let us begin by recalling that for any locally constant $\mathbb{Q}_l$-sheaf $\mathcal{G}$ on U,

$$H_c^0(\overline{U};\mathcal{G}) = 0 \quad \text{if } i=0 \text{ and } U \text{ is affine;}$$
$$H_c^2(\overline{U};\mathcal{G}) = (\mathcal{G}_u)_{\pi_1(\overline{U},u)} \otimes_{\mathbb{Q}_l} \mathbb{Q}_l(-1)$$
$$H_c^i(\overline{U};\mathcal{G}) = 0 \quad \text{if } i>2.$$

This gives that

$$L_{U,\mathcal{F}}(t) = \det(id - tF^*|_{H_c^1(\overline{U};\mathcal{F})})$$

and also that

$$L_{U,\mathcal{F}^{\otimes(2k)}}(t) = \frac{\det(id - tF^*|_{H_c^1(\overline{U};\mathcal{F}^{\otimes(2k)})})}{(1-tq^{k\beta+1})}.$$

In particular, note that the radius of convergence of $L_{U,\mathcal{F}}(t)$ is infinite, whereas that of $L_{U,\mathcal{F}^{\otimes(2k)}}(t)$ is $q^{-k\beta-1}$. The reader will see in a moment why we are only interested in the even tensor powers of $\mathcal{F}$.

Now let us turn to the local product description for $L_{U,\mathcal{F}^{\otimes(2k)}}(t)$.

Lemma 4.2. *The radius of convergence of each of the local factors*

$$\det(id - t^{\deg(x)}F_x^*|\mathcal{F}_x^{\otimes(2k)})^{-1}$$

is greater than or equal to the radius of convergence for $L_{\mathcal{U},\mathcal{F}^{\otimes(2k)}}(t)$.

Proof. Write $L_x(t)$ for the local factor at x, and write $L(t) = \prod_{x\in U} L_x(t)$ for the L-function. First observe that the logartihmic derivatives $L_x'(t)/L_x(t)$ are rational power series with positive coefficients. This follows from the formula

$$L_x'(t)/L_x(t) = \sum_{n=0}^{\infty} \mathrm{Tr}(F_x^n|\mathcal{F}^{\otimes(2k)})\cdot t^{n+1}$$

and the fact that $\mathrm{Tr}(F_x^n|\mathcal{F}^{\otimes(2k)}) = \mathrm{Tr}(F_x^n,\mathcal{F})^{2k}$. Here we are using the even powers!

It follows that $\log(L_x(t))$ is a power series with positive rational coefficients, as $\frac{d}{dt}[\log(L_x(t))] = L_x'(t)/L_x(t)$. Therefore $L_x(t) = \exp(\log(L_x(t)))$ also has positive rational coefficients.

Let $L(t) = \sum_n a_n t^n$ and $L_x(t) = \sum_n b_n t^n$. Then the product formula $L(t) = \prod_x L_x(t)$ and the positivity of the b's shows that $b_n \le a_n$ for all n. It follows that if $L(t)$ absolutely converges for some chosen t, then $L_x(t)$ also converges absolutely. This is what we wanted. □

Let α be an eigenvalue for F_x^* acting on $\mathcal{F}_x$. Then α^{2k} is an eigenvalue for F_x^* acting on $\mathcal{F}^{\otimes(2k)}(x)$, and so $\alpha^{-2k/\deg(x)}$ is a pole for the local factor at x. Therefore we must have

$$|\alpha^{-2k/\deg(x)}| \le q^{-k\beta-1},$$

which we may rewrite as

$$[q^{\deg(x)}]^{\frac{\beta}{2}+\frac{1}{2k}} \le |\alpha|.$$

As this holds for all $k\ge 1$, we therefore have

$$q_x^{\beta/2} \le |\alpha|.$$

But duality guarantees that if α is in eigenvalue of F_x^* acting on $\mathcal{F}_x$, then so is q_x^β/α. Applying the same argument as above to this second eigenvalue, we get

$$q_x^{\beta/2} \le |q_x^\beta/\alpha|,$$

or

$$|\alpha| \le q_x^{\beta/2}.$$

So $|\alpha| = q_x^{\beta/2}$, and this completes the proof of Proposition 4.1(a).

To prove Proposition 4.1(b) we return to the L-function for $\mathcal{F}$ itself:

$$L_{U,\mathcal{F}}(t) = \prod_{x\in|U|} \det(id - t^{\deg(x)} F_x|\mathcal{F}_x)^{-1} = \det(id - tF^*|_{H_c^1(\overline{U};\mathcal{F})}).$$

We know at this point that the radius of convergence for each local factor is $q^{-\beta/2}$. Using the fact that the points of U, indexing the product, are easy to understand, we will prove that the radius of convergence of the Euler product is at least $q^{-1-\beta/2}$. This implies, in particular, that the Euler product is not zero in this range. So if α is an eigenvalue for F acting on $H_c^1(\overline{U};\mathcal{F})$, then $\frac{1}{\alpha}$ is a zero of the L-function $L_{U,\mathcal{F}}(t)$, and therefore we must have

$$\left|\tfrac{1}{\alpha}\right| \ge q^{-1-\beta/2}.$$

Taking reciprocals we get $|\alpha| \le q^{1+\beta/2}$, as desired.

So what is left to do is to prove that the radius of convergence for the Euler product is at least $q^{-1-\beta/2}$ as claimed. Let N denote the rank of $\mathcal{F}$, and write

$$\det(id - tF_x|\mathcal{F}_x) = (1-\alpha_{x,1}t)(1-\alpha_{x,2}t)\cdots(1-\alpha_{x,N}t).$$

Then taking log of the Euler product gives

$$-\sum_{x\in U}\sum_{i=1}^{N} \log(1-\alpha_{x,i}t^{\deg(x)}),$$

and we must decide when this series converges. By a standard result from complex analysis (see [**A**, Theorem 6 of Chapter 5 and the discussion preceding it]), this series converges absolutely if and only if the series

$$\sum_{x\in U}\sum_{i=1}^{N} |\alpha_{x,i}t^{\deg(x)}|$$

converges absolutely. But now we are in business, as we know that $|\alpha_{x,i}| = q_x^{\beta/2}$ for each x and i. So we are really looking at the series

$$S(t) = N\cdot \sum_{x\in U} q_x^{\beta/2}|t^{\deg(x)}|.$$

If $|t| < q^{-1-\beta/2}$ then we can write

$$|t| = q^{-1-\beta/2}\cdot q^{-\epsilon}$$

for some $\epsilon > 0$. Then

$$S(t) < N\cdot \sum_{x\in U} q_x^{\beta/2}\cdot q_x^{-1-\beta/2-\epsilon} = N\sum_{x\in U} q_x^{-1-\epsilon} = N\cdot \sum_{i=0}^{\infty} u_i q^{(-1-\epsilon)i},$$

where u_i is the number of closed points x of U having $\deg(x) = i$.

But U is a subvariety of $\mathbb{A}^1$, and $\mathbb{A}^1$ has at most q^i points of degree i. So

$$S(t) < N\cdot \sum_{i=0}^{\infty} q^i\cdot q^{(-1-\epsilon)i} = \sum_{i=0}^{\infty}(q^{-\epsilon})^i,$$

and the series on the right converges because $q^{-\epsilon} < 0$.

5. Completion of the proof

Recall where we left off at the end of Section 2: we need to prove that the eigenvalues of F on $H^1_c(\overline{U}; \mathcal{E})$ have absolute norm $q^{d/2}$. The sheaf $\mathcal{E}$ has a skew-symmetric bilinear form $\psi\colon \mathcal{E} \otimes \mathcal{E} \to \mathbb{Q}_l(-d/2)$, but we do not know that this pairing is nondegenerate (this is the content of the Hard Lefschetz Theorem!) The proof now breaks down into three cases:

Case 1: $\mathcal{E} \cap \mathcal{E}^{\perp} = \mathbf{0}$, i.e., the pairing ψ is nondegenerate.

This is the main case. In fact, if the Hard Lefschetz Theorem is true then this *must* happen; but we don't know the Hard Lefschetz Theorem yet.

The cup product on $H^{d-1}(Y; \mathbb{Q}_l)$ gives rise to an alternating bilinear form $\mathcal{E} \otimes \mathcal{E} \to \mathbb{Q}_l(-d/2)$. Since $\mathcal{E} \cap \mathcal{E}^{\perp} = 0$, this form is nondegenerate. In this case we refer to the "Fundamental Estimate", Proposition 4.1(b), and we are done.

Case 2: ????.

Case 3: $\mathcal{E} \subseteq \mathcal{E}^{\perp}$**.** ???

Part 3

Algebraic K-theory

CHAPTER 8

Algebraic K-theory

The subject of K-theory spans both algebraic topology and algebraic geometry, and has a circuitous history. It starts in algebraic geometry, with Grothendieck's introduction of the group $K(X)$—for X a scheme—in his work on the Riemann-Roch theorem. It them jumps to algebraic topology, where Atiyah and Hirzebruch took the analog of $K(X)$—now for X a topological space—and extended it to an entire cohomology theory $K^*(X)$. Grothendieck's original group is now written $K^0(X)$. Back in algebraic geometry, there were then efforts over about ten years to define "algebraic" versions of the Atiyah-Hirzebruch groups $K^*(X)$, where X is once again a scheme. This culminated with the work of Quillen in the early 1970s, giving us a definitive version of what one now calls "higher algebraic K-theory".

In this chapter our aim is to give a brief overview of this subject. We start at the very beginning, with the connection between vector bundles and projective modules.

There is an unfortunate notational annoyance which comes up when dealing with algebraic K-theory, and we might as well get this out of the way at the beginning. In algebraic K-theory, the group which is the "analog" of the topological group $K^n(X)$ is unfortunately written $K_{-n}(X)$. This is partly because these groups turn out to be the most interesting when n is *negative*, and so the algebraic K-theory notation eliminates having to write a bunch of minus signs everywhere. But the notation is very unfortunate, because something which is trying hard to be a cohomology theory ends up not really "looking" like a cohomology theory. It is also rather annoying, when X is a variety over the complex numbers, to have to write the comparison map from algebraic to topological K-theory as

$$K_n(X)_{alg} \longrightarrow K^{-n}(X)_{top}.$$

Throughout this chapter we will constantly mix notations, and write

$$K_n(X) = K^{-n}(X)$$

for the same group. Our preference, thought, will always be for the latter, cohomological notation, because this results in the most natural-looking formulas. The former notation is in some sense forced on us because it is what everyone uses, but it would be better if it could be abandoned altogether.

??????

Part 4

Motives and other topics

CHAPTER 9

Motives

In this chapter we attempt to give a first look at motives. Our discussion will often be a bit wishy-washy. The aim is not to give a careful, rigorous exposition but rather to give an overall picture for how one can think about this material.

1. Topological motives

There is indeed a theory of "motives" for topological spaces, although one does not usually use that term. We'll start by talking about this because it will be a very useful guide when we start fantasizing about motives for algebraic varieties.

Let $\mathcal{A}b(\mathcal{T}op)$ denote the category of abelian group objects in topological spaces—that is, the category of topological abelian groups. We will sometimes write this as $\mathcal{A}b\mathcal{T}op$ when the parentheses become cumbersome. There are some evident objects in this category: the discrete abelian groups, the tori $S^1 \times \cdots \times S^1$, and extensions of tori by discrete abelian groups. Every topological abelian group which is a finite CW-complex is such an extension of a torus. To see this, note that in an n-dimensional CW-complex there are points with neighborhoods homeomorphic to $\mathbb{R}^n$. In a topological group, every point looks like every other point locally—so if the group is a finite CW-complex, it is actually a manifold. Some nontrivial smoothing theory shows that it is therefore a compact abelian Lie group, and the classification of these is well-known.

It is important to realize that $\mathcal{A}b(\mathcal{T}op)$ has many objects besides these finite-dimensional ones. Let X be any topological space. The free abelian group on the underlying set of X inherits a topology from X, and we will denote this new space as $\mathcal{A}b(X)$. The resulting functor $\mathcal{A}b\colon \mathcal{T}op \to \mathcal{A}b(\mathcal{T}op)$ is the left adjoint to the forgetful functor. If (X, x) is a pointed space it is useful to also consider the reduced abelianization $\widetilde{\mathcal{A}b}(X) = \mathcal{A}b(X)/\langle x\rangle$. Here we are just forcing x to be the zero element of our group.

The spaces $\mathcal{A}b(X)$ give us a multitude of objects in $\mathcal{A}b(\mathcal{T}op)$. They are typically infinite-dimensional, however. Here are a few examples worth mentioning:

Example 1.1. $\widetilde{\mathcal{A}b}(S^1) \cong S^1$, $\widetilde{\mathcal{A}b}(\mathbb{R}P^2) \cong \mathbb{R}P^\infty$, and $\widetilde{\mathcal{A}b}(S^2) \cong \mathbb{C}P^\infty$. We explain the last isomorphism, and leave the reader to ponder the others. Regard S^2 as the Riemann sphere $\mathbb{C}P^1$, or better yet the extended complex plane $\mathbb{C} \cup \infty$. Take ∞ to be the basepoint. Regard $\mathbb{C}P^\infty$ as the space of lines in the infinite-dimensional vector space $\mathbb{C}(t)$.

An element of $\widetilde{\mathcal{A}b}(S^2)$ is a formal word $\sum_i n_i[z_i]$. Define a map $h\colon \widetilde{\mathcal{A}b}(S^2) \to \mathbb{C}P^\infty$ by sending $\sum_i n_i[z_i]$ to the line in $\mathbb{C}[t]$ spanned by the rational function $\prod_i (t - z_i)^{n_i}$. This is the rational function whose set of zeros *is* the formal sum $\sum_i n_i[z_i]$, where we are counting zeros with multiplicities and regarding poles as zeros with negative multiplicity. It is easy to see that h is continuous and injective.

The fundamental theorem of algebra shows that h is surjective. We leave it to the reader to complete the proof that h is a homeomorphism.

The following is the most important theorem about abelianizations:

THEOREM 1.2 (Dold-Thom). *For any space X one has canonical isomorphisms $\pi_i(\mathcal{A}b(X), 0) \cong H_i(X;\mathbb{Z})$. If X is pointed then there are canonical isomorphisms $\pi_i\widetilde{\mathcal{A}b}(X) \cong \tilde{H}_i(X;\mathbb{Z})$.*

Note that by the Dold-Thom theorem it follows that $\widetilde{\mathcal{A}b}(S^n)$ is a $K(\mathbb{Z}, n)$. More generally, if $n \geq 1$ and $M(n, q)$ denotes the cofiber of the multiplication-by-q map $S^n \to S^n$ then $\widetilde{\mathcal{A}b}(M(n, q))$ is a $K(\mathbb{Z}/q, n)$.

The category $\mathcal{A}b(\mathcal{T}op)$ has a model category structure in which a map is a weak equivalence or fibration if and only if it is so when regarded in $\mathcal{T}op$. The functor $\mathcal{A}b$ is a left Quillen functor, and so for X a cofibrant space and $Z \in \mathcal{A}b(\mathcal{T}op)$ one has the adjunction formula

$$\mathrm{Ho}\,(\mathcal{T}op)(X, Z) \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathcal{A}b(X), Z). \tag{1.3}$$

For pointed spaces X one has

$$\mathrm{Ho}\,(\mathcal{T}op_*)(X, Z) \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\widetilde{\mathcal{A}b}(X), Z)$$

where the basepoint of Z is its zero element. Using that $\widetilde{\mathcal{A}b}(S^n)$ is a $K(\mathbb{Z}, n)$, we can now write formulas such as

$$H^n(X;\mathbb{Z}) \cong [X, K(\mathbb{Z}, n)] \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathcal{A}b(X), \widetilde{\mathcal{A}b}(S^n)).$$

We can also represent singular homology entirely within the context of $\mathcal{A}b(\mathcal{T}op)$, via the formula

$$H_n(X;\mathbb{Z}) \cong \mathrm{Ho}\,(\mathcal{T}op_*)(S^n, \mathcal{A}b(X)) \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\widetilde{\mathcal{A}b}(S^n), \mathcal{A}b(X)).$$

The point is that singular homology and cohomology are representable in $\mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)$. Of course they are also representable in $\mathcal{T}op$, but $\mathcal{T}op$ has quite a bit of information which can't be seen by singular cohomology. We will see below that $\mathcal{A}b(\mathcal{T}op)$ *only* has information which can be seen by singular cohomology.

1.4. Algebraic models. As a model category, $\mathcal{T}op$ is Quillen equivalent to $s\mathcal{S}et$. One can then see that $\mathcal{A}b(\mathcal{T}op)$ is Quillen equivalent to $\mathcal{A}b(s\mathcal{S}et)$. Note that the latter is just the category of simplicial abelian groups, with its usual model structure where weak equivalences and fibrations are determined by forgetting into $s\mathcal{S}et$. Finally, recall that the category of simplicial abelian groups is equivalent to the category of non-negatively graded chain complexes of abelian groups. So we have Quillen equivalences

$$\mathcal{A}b(\mathcal{T}op) \simeq \mathcal{A}b(s\mathcal{S}et) \simeq \mathrm{Ch}_{\geq 0}(\mathbb{Z}).$$

This gives an algebraic model for $\mathcal{A}b(\mathcal{T}op)$. What chain complex does $\mathcal{A}b(X)$ correspond to under these maps? It is just the singular chain complex of X.

1.5. Splittings. Every space X can be constructed up to homotopy as a cell complex. A related fact is that every topological abelian group decomposes—up to homotopy—as a product of Eilenberg-MacLane spaces. Probably the easiest way to prove this is to use the Quillen equivalence of $\mathcal{A}b(\mathcal{T}op)$ with $Ch_{\geq 0}(\mathbb{Z})$ to reduce the problem to homological algebra. It is a well-known fact that every non-negatively graded chain complex is quasi-isomorphic to its homology, regarded as a chain complex with zero differential.

This observation has an important consequence. If X and Y are any two spaces, then one can consider the abelian groups

$$\mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathcal{A}b(X), \mathcal{A}b(Y)).$$

But since $\mathcal{A}b(Y)$ splits as a product of Eilenberg-MacLane spaces, this abelian group splits into products of singular cohomology groups of X, with various shiftings and coefficients. Thus, studying the category $\mathcal{A}b(\mathcal{T}op)$ is really just studying singular cohomology.

1.6. Motivic notation. The space $\mathcal{A}b(X)$ should be called the **unstable topological motive** of X, and we'll sometimes denote it $M(X)$. The space $\widetilde{\mathcal{A}b}(X)$ is the **unstable reduced motive**, denoted $\widetilde{M}(X)$. The space $\widetilde{\mathcal{A}b}(S^n)$ is denoted $\mathbb{Z}[n]$. Under the Quillen equivalence with chain complexes, $\mathbb{Z}[n]$ is just a chain complex with $\mathbb{Z}$ in dimension n and zeros elsewhere.

The category $\mathcal{A}b(\mathcal{T}op)$ is the category of "unstable topological motives", and its homotopy category is the "derived category of (unstable) topological motives".

Note that $\mathcal{A}b(\mathcal{T}op)$ has a tensor product: if X and Y are topological abelian groups then the algebraic tensor product inherits a topology in a natural way. Tensoring with $\mathbb{Z}[1]$ is just the suspension in the model category $\mathcal{A}b(\mathcal{T}op)$ (use the equivalence with chain complexes, for instance). One sometimes writes $X[k]$ for $X \otimes \mathbb{Z}[k]$.

In our new notation our formulas for singular homology and cohomology become

$$H^n(X;\mathbb{Z}) \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathcal{A}b(X), \mathbb{Z}[n])$$

and

$$H_n(X;\mathbb{Z}) \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathbb{Z}[n], \mathcal{A}b(X)).$$

1.7. Stable motives. One can stabilize the category of spaces to form spectra. One can stabilize the category of non-negatively graded chain complexes to get all $\mathbb{Z}$-graded chain complexes. In the same way, we can stabilize the model category $\mathcal{A}b(\mathcal{T}op)$. This produces a model category we'll call the category of stable topological motives. It is Quillen equivalent to $\mathbb{Z}$-graded chain complexes. The suspension spectrum of $\mathcal{A}b(X)$ will be called the (stable) topological motive of X, and will also be denoted $M(X)$ by abuse.

Note that the category $\mathcal{A}b(\mathcal{T}op)$ was already semi-stable, in the sense that

$$\mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(A, B) \cong \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\Sigma A, \Sigma B).$$

Here Σ denotes the suspension in $\mathcal{A}b(\mathcal{T}op)$, which is *not* the suspension of underlying topological spaces; the suspension in $\mathcal{A}b(\mathcal{T}op)$ is instead tensoring with $\widetilde{\mathcal{A}b}(S^1)$. In motivic notation, it is $A \mapsto A[1]$. Here again, it is probably easiest to justify the above semi-stability formula by working in the category $\mathrm{Ch}_{\geq 0}(\mathbb{Z})$.

Because of this semi-stability, the process of stabilization does not really do much to the maps...it only adds new objects, namely formal desuspensions and colimits of such things. So for instance, if $A, B \in \mathcal{A}b(\mathcal{T}op)$ then

$$\mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(A,B) \cong \mathrm{Ho}\,(\mathcal{S}pectra(\mathcal{A}b\mathcal{T}op))(\Sigma^\infty A, \Sigma^\infty B).$$

Again, compare the passage from $\mathrm{Ch}_{\geq 0}(\mathbb{Z})$ to $\mathrm{Ch}(\mathbb{Z})$.

1.8. The moral. There are a multitude of generalized cohomology theories for topological spaces, but among these one can isolate certain special ones which we'll call "abelian". These are the cohomology theories which can be calculated as the homology groups of a chain complex functorially associated (contravariantly) to every space X. So these include singular cohomology theory with any coefficients and shifts of such cohomology theories, but excludes things like K-theory and cobordism.

One way to obtain an abelian cohomology theory is to start with a spectrum object E in $\mathcal{A}b(\mathcal{T}op)$. One gets an ordinary spectrum by forgetting the abelian group structures. For each X, the mapping space (or spectrum) from $\mathcal{A}b(X)$ to E can be modelled by a chain complex. So the E-cohomology groups of X are indeed the cohomology groups of functorial chain complexes.

By adapting the classical proof of Brown representability, one can in fact show that every abelian cohomology theory is represented by a spectrum object over $\mathcal{A}b(\mathcal{T}op)$—that is, by an element of $\mathcal{S}pectra(\mathcal{A}b\mathcal{T}op)$. So this category is in some sense encapsulating everything that can be studied using abelian cohomology theories.

One can think of the stable motive $M(X)$ of a space as a 'shadow' of X, in which many of the features of X have been lost—the things that haven't been lost are presicely those things which can be seen by some abelian cohomology theory.

This perspective may seem strange when applied to topological spaces, because all abelian cohomology theories break up into pieces that look like singular cohomology theory with coefficients. The category of stable motives is just the category of chain complexes, and all we are saying is that the singular chain complex of X encodes exactly that information which can be seen by singular cohomology. It is practically a tautology. The real point of this perspective is not what it gives us for spaces, but what it may give us in the setting of other model categories.

The picture we've described in this section—that of looking at the abelian group objects in $\mathcal{T}op$ and using this category as a kind universal setting for singular homology cohomology—was introduced into topology by Quillen. He used it to define what is now called André-Quillen cohomology of commutative rings, by applying the same ideas to the model category of simplicial commutative rings. Quillen was surely very much aware of Grothendieck's ideas about motives, so it seems likely they influenced him here.

2. Motives for algebraic varieties

Let k be a field and consider the category Sm/k of smooth schemes of finite type over k. More generally one might consider all schemes of finite type, but for the moment let's be content with smooth ones.

To each scheme X one can associated the sequence of l-adic cohomology groups $H^*(X; \mathbb{Z}_l)$, one sequence for every prime l different from the characteristic of k. If X is defined over $\mathbb{C}$ one can also look at the singular cohomology groups of the analytic space $X(\mathbb{C})$. There are also crystalline cohomology groups, algebraic de Rham cohomology groups, and probably others. These cohomology theories have a number of things in common with each other.

Grothendieck envisioned a category $\mathcal{M}(k)$ which would be something like a 'shadow' of the category Sm/k. This category would *only* see cohomological information about smooth schemes. There would be a functor $M\colon Sm/k \to \mathcal{M}(k)$ which sent every scheme to its 'shadow image', and all the cohomology theories mentioned in the last paragraph would factor through this functor. Theorems which held for all of these cohomology theories were supposed to really be theorems about the category $\mathcal{M}(k)$. Each object $M(X)$ was called the **motive** of the scheme X, and $\mathcal{M}(k)$ was called the category of motives. Sometimes Grothendieck and his followers described the functor M as a 'universal cohomology theory'.

One thing we should mention right away is that when Grothendieck and others of that era talked about 'cohomology theories' for varieties they only had in mind analogs of *abelian* cohomology theories. Algebraic K-theory—as a full cohomology theory—hadn't even been invented yet, so they weren't thinking about things like that at all. Nowadays we have to be careful to say that one would not expect algebraic K-theory to factor through M, only those cohomology theories of a more 'abelian' nature.

Based on our discussion of topological motives from the last section (a set of ideas which was not available to Grothendieck and only later introduced by Quillen), we see now that Grothendieck was wanting something like a category of abelian group objects $\mathcal{A}b(Sm/k)$. Yet this is not quite right as stated. An abelian group object in Sm/k is just an abelian variety. While there are many of these, under our analogy with topology they are really only giving us those abelian group objects which are finite-dimensional—the torii and extensions of torii by discrete groups. For reasons we will see in just a moment, this is enough to basically construct $M(X)$ when X is an algebraic curve. But as soon as we move to varieties of dimension at least two, we would expect $M(X)$ *not* to just decompose into abelian varieties. The moral is that to construct $\mathcal{M}(k)$ we will need a host of new objects.

2.1. Getting familiar with motives. For the moment let us postpone any discussion of how to construct $\mathcal{M}(k)$ and instead focus on what we expect this category to be like. I want to develop some conventions about notation and also write down a series of 'facts' one would like to hold. The point for the moment is not to do any serious mathematics but rather to build up some intuition for how to work with motives.

(1) Based on the situation in topology, we can hope for a stable, additive model category $\mathcal{M}(k)$ and a functor $M\colon Sm/k \to \mathcal{M}(k)$.
(2) We also expect to have an embedding $j\colon \mathcal{A}b(Sm/k) \to \mathcal{M}(k)$—that is to say, every abelian variety should be a motive. If J is an abelian variety we will usually simplify $j(J)$ to just J.

    Discussion: The situation is a little confusing because on the one hand J *has* a motive $M(J)$, but at the same time J *is* a motive via $j(J)$. And $M(J)$ and $j(J)$ will almost always be different. The thing to remember is that $M(J)$ only depends on the underlying variety of J, not the group structure; whereas $j(J)$ really depends on the group structure.

(3) If X is an object in $\mathcal{M}(k)$ we will denote its suspension by either ΣX (topologists' notation) or $X[1]$ (geometers' notation).
(4) The object $M(\operatorname{Spec} k)$ will be very special, and we'll denote it by $\mathbb{Z}(0)$ (in analogy with the topological case, where the motive of a point is the discrete abelian group $\mathbb{Z}$). Sometimes we'll abbreviate $\mathbb{Z}(0)$ to just "$\mathbb{Z}$"; for instance, we will write $\mathbb{Z}[i]$ instead of $\mathbb{Z}(0)[i]$. But note that this is just notation—the object $\mathbb{Z}(0)$ is certainly not just the usual abelian group $\mathbb{Z}$.
(5) For every $X \in Sm/k$ we define the reduced motive $\widetilde{M}(X)$ to be the homotopy fiber of $M(X) \to M(\operatorname{Spec} k)$. Note that if X has a rational point then this map has a splitting, so that $M(X) \simeq \widetilde{M}(X) \oplus \mathbb{Z}(0)$.
(6) The category $\mathcal{M}(k)$ should have a tensor product which descends to the homotopy category. The unit will be $\mathbb{Z}(0)$.
(7) In addition to the object $\mathbb{Z}(0)$ just defined, there will be a special object $\mathbb{Z}(1)$ called the "Tate object". For $q \geq 0$ we define $\mathbb{Z}(q) = \mathbb{Z}(1)^{\otimes q}$. Also, for each $q < 0$ there will be a special object $\mathbb{Z}(q)$ and there will be isomorphisms

$$\mathbb{Z}(i) \otimes \mathbb{Z}(j) \cong \mathbb{Z}(i+j)$$

for all $i, j \in \mathbb{Z}$.

For $A \in \mathcal{M}(k)$ the object $A \otimes \mathbb{Z}(q)$ will be denoted $A(q)$.

Discussion: There isn't really a topological analog of the object $\mathbb{Z}(q)$. The best thing is to think about it as a 'twisted' version of $\mathbb{Z}(0)$.

We can explain things a little further as follows. Consider the scheme $\mathbb{A}^1 - 0$. If we were working over the complex numbers this would be a circle (up to homotopy), and its reduced motive would therefore also be a circle—i.e., the topological motive $\mathbb{Z}[1]$. In algebraic geometry the reduced motive of $\mathbb{A}^1 - 0$ will not just be $\mathbb{Z}[1]$, but it will be an object which has a similar importance in terms of cohomology in degree 1. The Tate object $\mathbb{Z}(1)$ is just the desuspension of $\widetilde{M}(\mathbb{A}^1 - 0)$—that is,

$$\mathbb{Z}(1) \cong \widetilde{M}(\mathbb{A}^1 - 0)[-1] \quad \text{or} \quad \widetilde{M}(\mathbb{A}^1 - 0) \cong \mathbb{Z}(1)[1].$$

(8) The objects $\mathbb{Z}(q)[n]$ will be called 'motivic Eilenberg-MacLane spaces'.
(9) The fact that $\mathcal{M}(k)$ is a stable model category shows that there will be isomorphisms

$$\operatorname{Ho}(\mathcal{M}(k))(A, B) \cong \operatorname{Ho}(\mathcal{M}(k)(A[1], B[1])$$

for any $A, B \in \mathcal{M}(k)$. Moreover, for any object Z the fact that tensoring with Z is an additive functor gives us maps

$$\operatorname{Ho}(\mathcal{M}(k))(A, B) \to \operatorname{Ho}(\mathcal{M}(k))(A \otimes Z, B \otimes Z).$$

If Z is invertible under the tensor product, then this map will be an isomorphism. So this applies in particular when Z is the object $\mathbb{Z}(q)$, giving us isomorphisms

$$\operatorname{Ho}(\mathcal{M}(k))(A, B) \cong \operatorname{Ho}(\mathcal{M}(k))(A(q), B(q))$$

for any $q \geq 0$.

Discussion: So $\mathcal{M}(k)$ is "stable in two directions".
(10) For any object $A \in \mathcal{M}(k)$ one obtains a bigraded sequence of functors on smooth schemes

$$X \mapsto A^{p,q}(X) = \operatorname{Ho}(\mathcal{M}(k))(M(X), A(q)[p]).$$

We'll call this the bigraded A-cohomology of X.

The **motivic cohomology groups** of X are the bigraded groups

$$H^{p,q}(X;\mathbb{Z}) = \mathrm{Ho}\,(\mathcal{M}(k))(M(X), \mathbb{Z}(q)[p]).$$

The **motivic homology groups** of X are the groups

$$H_{p,q}(X;\mathbb{Z}) = \mathrm{Ho}\,(\mathbb{Z}(q)[p], M(X)).$$

These groups are more commonly denoted

$$H^{p,q}(X;\mathbb{Z}) = H^p(X;\mathbb{Z}(q)) \qquad \text{and} \qquad H_{p,q}(X;\mathbb{Z}) = H_p(X;\mathbb{Z}(q)).$$

We will find ourselves going back and forth between the two notations indiscriminately.

(11) If $A \in \mathcal{M}(k)$ it will be convenient for us to introduce the notation $\pi_i(A) = \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathbb{Z}[i], A)$ and more generally

$$\pi_{p,q}(A) = \mathrm{Ho}\,(\mathcal{A}b\mathcal{T}op)(\mathbb{Z}(q)[p], A).$$

(12) If $X \in Sm/k$ then the projection map $X \times \mathbb{A}^1 \to X$ should induce a weak equivalence $M(X \times \mathbb{A}^1) \simeq M(X)$.

(13) If $\{U, V\}$ is a Zariski open cover for a smooth scheme X then there should be a homotopy cofiber sequence of the form

$$M(U \cap V) \to M(U) \oplus M(V) \to M(X) \to \Sigma M(U \cap V).$$

Discussion: Note that applying $\mathrm{Ho}\,(\mathcal{M}(k))(-, \Sigma^* A(q))$ will then induce a long exact Mayer-Vietoris sequence for $A^{*,q}(-)$. In particular, we get a Mayer-Vietoris sequence for motivic cohomology.

(14) If $E \to X$ is an algebraic vector bundle of rank n then

$$M(\mathbb{P}(E)) \simeq M(X) \oplus M(X)(1)[2] \oplus M(X)(2)[4] \cdots \oplus M(X)(n)[2n].$$

Written differently, $M(E) \simeq \oplus_{i=0}^{n} M(X) \otimes L^n$ where $L = \mathbb{Z}(1)[2]$.

(15) If X is a smooth projective curve then there will be a splitting

$$M(X) \simeq \mathbb{Z}(0) \oplus \mathrm{Jac}(X) \oplus \mathbb{Z}(1)[2]$$

where $\mathrm{Jac}(X)$ is the Jacobian variety of X (which is an abelian variety, regarded as a motive via the functor j).

Discussion: For this one, we should compare the analogous situation in topology. Let W be a compact Riemann surface. Then the topological motive of W splits as a product of Eilenberg-MacLane spaces, and the types can be predicted from $H_*(W)$ using the Dold-Thom theorem:

$$M_{top}(W) \simeq \mathbb{Z} \times [K(\mathbb{Z}, 1) \times \cdots K(\mathbb{Z}, 1)] \times K(\mathbb{Z}, 2).$$

Here there are $2g$ copies of $K(\mathbb{Z}, 1)$, where g is the genus of W. Note that the Jacobian variety of W is topologically a torus, and it is known classically that the rank is $2g$. So the product of $K(\mathbb{Z}, 1)$'s in the above splitting is precisely the Jacobian variety.

(16) If X is a smooth projective scheme of dimension d then there will be a splitting

$$M(X) \simeq \mathbb{Z}(0) \oplus e(X) \oplus \mathbb{Z}(d)[2d].$$

(17) It was conjectured by Grothendieck that any $M(X)$ should split up as

$$M(X) \simeq h_0(X) \oplus h_1(X) \oplus h_2(X) \oplus \cdots h_d(X)$$

where d is the dimension of X, and where $h_i(X)$ is something like the "motive of X related to i-dimensional cohomology". Note that such a splitting also happens in the topological world.

Grothendieck could obtain this splitting if he knew that the Künneth components of the diagonal $\Delta \in H^*_{et}(X \times X; \mathbb{Z}_l) \otimes \mathbb{Q}_l$ were all algebraic. This remains an open problem.

(18) In the topological setting every object was built, up to homotopy, from spheres, and it resulted that every motive split as a product of Eilenberg-MacLane spaces. So when computing maps from a motive A into a motive B, one could typically split B into factors looking like $\mathbb{Z}[n]$'s and $\mathbb{Z}/q[m]$'s; it followed that the hom sets in the category of motives all just boiled down to singular cohomology.

In the algebraic geometry world, it does not seem to be possible to construct all varieties from some simple building blocks. So while one still has the special groups $\mathrm{Ho}\,(\mathcal{M}(k))(A, \mathbb{Z}(q)[p])$ and $\mathrm{Ho}\,(\mathcal{M}(k))(A, \mathbb{Z}/n(q)[p])$, these are probably not enough to understand $\mathrm{Ho}\,(\mathcal{M}(k))(A, B)$ for all motives B. In particular, one can study the sets $\mathrm{Ho}\,(\mathcal{M}(k))(M(X), M(Y)(q)[p])$ for smooth schemes X and Y. Such groups were explored, for instance, in a paper of Friedlander and Voevodsky [**FV**] under the name "bivariant cycle cohomology".

2.2. Comparison with the classical picture of motives. The picture of motives we have developed differs in some ways from the one Grothendieck originally painted.

(1) Grothendieck wanted $\mathcal{M}(k)$ to be an abelian category, and he did not consider anything like an associated homotopy category. Beilinson seems to be the first one to talk about a derived category of motives, and this is what was eventually constructed by Voevodsky. Constructing an *abelian* category of motives seems to be elusive. Whether or not it can be done has not been relevant for anything we have considered here, but apparently it is important in relation to certain conjectures on algebraic cycles.

(2) One can speak of 'homological motives' and 'cohomological motives'. As an analogy, consider the difference between the covariant functor $\mathcal{T}op \to \mathrm{Ch}(\mathbb{Z})$ which sends any space X to its singular chain complex $S_*(X)$, as opposed to the contravariant functor $\mathcal{T}op \to \mathrm{Ch}(\mathbb{Z})$ which sends any space to its singular *cochain* complex $\mathrm{Hom}(S_*(X), \mathbb{Z})$. Certainly these two functors have the same information in them, but it's organized slightly differently in the two cases.

Our discussion has always been about 'homological' motives, because it is in this context that the topological analogies really become clear. In classical algebraic geometry people always talked about cohomological motives, however. All the ideas are basically the same, but the variance on everything is reversed.

(3) The classical work on motives was all done *rationally*. So the categories of motives were not just additive categories, they were categories where the hom sets were rational vector spaces. This was necessary in part because the Künneth isomorphism was needed in order to get the ad hoc constructions of motives off the ground.

3. Constructing categories of motives

Grothendieck constructed a category of "pure" motives, sometimes called Chow motives. The Standard Conjectures on algebraic cycles implied certain properties of this category, but as of yet these Standard Conjectures remain unproven.

From the point of view of a topologist, the category of Chow motives seems somewhat crude and deficient. Voevodsky constructed a model category (although he didn't quite say it that way) whose homotopy category was to be the derived category of motives. This requires much more machinery than Grothendieck's construction, but in the end it gives a more satisfying result.

3.1. Chow motives. Let's imagine how one might construct the category $\mathcal{M}(k)$. First, for any $X \in Sm/k$ one will need an object $\bar{X} \in \mathcal{M}(k)$ to serve as its motive $M(X)$. To understand the maps from $\bar{X}$ to $\bar{Y}$ we proceed in analogy with the topological case. If W and Z are spaces, maps of abelian groups $\mathcal{A}b(W) \to \mathcal{A}b(Z)$ are in bijective correspondence with continuous maps $W \to \mathcal{A}b(Z)$. Such a map assigns to each point $x \in W$ a formal um $f(w) = \sum_i n_i[z_i]$. We can imagine the "graph" of such a thing in $W \times Z$, and it sort of looks like a certain kind of cycle. We get something like a formal linear combination of subspaces of $W \times Z$, where each of these subspaces is a branched cover of W via the projection map $W \times Z \to W$.

Lefschetz had long ago considered the concept of a *correspondence*. Every map $f\colon W \to Z$ gives a cycle on $W \times Z$ by taking its graph, and Lefschetz realized that more general cycles on $W \times Z$ could be thought of as 'generalized maps'. In particular, they can be composed and one can use them to get induced maps on homology. This same idea can be used in algebraic geometry.

Let X and Y be two objects in Sm/k. A **correspondence** from X to Y is an element of $\mathrm{CH}^*(X \times Y)$. If X is connected, a **degree zero** correspondence is one of dimension $\dim X$ (if we were constructing cohomological motives we would require that the codimension of the cycle is $\dim X$). More generally, a degree zero correspondence is one which arises from a cycle of dimension $\dim U$ on every component U of X. The degree zero correspondences from X to Y play the role of our "cycles on $X \times Y$ made up of branched covers".

We can make a first approximation to $\mathcal{M}(k)$ by saying that the objects are formal symbols $\bar{X}$, one for every $X \in Sm/k$, and that the maps from $\bar{X}$ to $\bar{Y}$ are the set of degree zero correspondences from X to Y. Composition comes from the composition of correspondences. One calls this the **correspondence category**.

The problem with the correspondence category is that it is not very robust. It is an additive category, but it typically does not have kernels, images, or cokernels. One can begin to fix this in the following way. Let $\mathcal{C}$ be an additive category. A projector in $\mathcal{C}$ is a map $f\colon X \to X$ such that $f^2 = f$. One can formally add an image for all projectors. Define a category $\hat{\mathcal{C}}$ whose objects are pair (X, f) where $X \in \mathcal{C}$ and f is a projector. Think of this as the formal image of f. The abelian group of maps from (X, f) to (Y, g) is the group of all $\alpha\colon X \to Y$ such that $f\alpha = \alpha g$ modulo the subgroup of those α such that $f\alpha = 0$. Note that there is a full and faithful embedding $\mathcal{C} \to \hat{\mathcal{C}}$ sending X to (X, id).

Applying the above hat-construction to the category of correspondences gives the category $\mathcal{M}_{CH}(k)$ of **effective Chow motives**.

Some missing stuff here...still need to invert the Tate motive to get the full category of motives. Yuck.

REMARK 3.2 (A clarification). Given any reasonable cohomology theory for smooth schemes, Grothendieck showed how to construct a category of motives based on that cohomology theory. The above construction is the one based on the rational Chow groups $\mathrm{CH}^*(-)\otimes\mathbb{Q}$, and is therefore referred to as the category of Chow motives. This seems to be the most interesting choice, as any reasonable cohomology theory will receive a map from the Chow groups via the fundamental classes of algebraic cycles. So the choice to use Chow groups is an attempt to give the most generic construction.

Grothendieck was also very interested in the corresponding construction where the Chow groups were replaced by the rational vector spaces of cycles modulo numerical equivalence. Often when people refer to "pure motives" they mean the category of motives constructed from this theory.

3.3. Voevodsky's motives.

4. Motivic cohomology and spaces of algebraic cycles

Let us go back to the topological case for some intuition. For a space X we looked at the topological abelian group $\mathcal{A}b(X)$ and found that its homotopy groups were the singular homology groups of X. Note that $\mathcal{A}b(X)$ can be regarded as the space of 0-cycles on X. Is there a space of n-cycles for any n?

Probably with some effort one could define a topological abelian group $Z_n(X)$ which could reasonably be called the space of n-cycles on X, but no one has really done this. The reason is that one expects a weak equivalence in $\mathcal{A}b\mathcal{T}op$ of the form

$$Z_n(X) \simeq \Omega^n Z_0(X).$$

So the information in $Z_n(X)$ is really just the same information as in $Z_0(X)$, but shifted.

In fact, here is one crude construction of $Z_n(X)$. Start with the simplicial abelian group of singular chains on X, loop it down n times by your favorite looping maching, and then construct the geometric realization. This is kind of silly, though.

Another construction of $Z_n(X)$ was given by Almgren in the case where X is a complex algebraic variety. He defined a space of analytic n-cycles on X, and he proved exactly that $Z_n(X)$ was homotopy equivalent to $\Omega^n Z_0(X)$.

Okay, so considering higher dimensional cycles doesn't really get us anything new in the topological world. But in the world of algebraic geometry it *does* give something new. Let us postulate that in our category of motives $\mathcal{M}(k)$ we have not only the object $M(X) = Z_0(X)$ for any $X \in Sm/k$ but also objects $Z_r(X)$ for any $r \geq 0$, representing spaces of algebraic cycles of dimension r. Here, the formula we want turns out to be

$$Z_r(X) \simeq \Omega^{2r,r} Z_0(X) \simeq Z_0(X)(-r)[-2r].$$

Or more generally,

$$Z_{r+1}(X) \simeq \Omega^{2,1} Z_r(X) \simeq Z_n(X)(-1)[-2].$$

Playing with the Dold-Thom theorem just at a formal level, we find that we expect

$$\pi_i Z_r(X) = \pi_{2r+i,r} Z_0(X) = H_{2r+i,r}(X).$$

If X is smooth and projective of dimension d then we can use Poincaré Duality to write

$$H_{2r+i,r}(X) \cong H^{2d-2r-i,d-r}(X).$$

We can use these ideas to understand Bloch's higher Chow groups. For any smooth scheme X and any $p \geq 0$ Bloch constructed a complex $\mathrm{CH}^p(X,-)$ which should be thought of as an algebraic model for the "space of codimension p cycles on X". He defined

$$\mathrm{CH}^p(X,q) = H_q(CH^p(X,-)).$$

If we compare this to the above we see that we are looking at

$$\mathrm{CH}^p(X,q) = \pi_q Z_{d-p}(X) \cong H_{2d-2p+q,d-p}(X) \cong H^{2p-q,p}(X).$$

This formula gives the translation from higher Chow groups to motivic cohomology.

We can use these ideas to understand other constructions of motivic cohomology as well. For instance, in ???? Voevodsky considers an object analogous to the following. Start with $W = M((\mathbb{A}^1-0)\times\cdots M(\mathbb{A}^1-0))$ (n factors) and let J be the sums of all the 'images' of the maps

$$j_i\colon M((\mathbb{A}^1-0)\times(\mathbb{A}^1-0)\times\cdots\times(\operatorname{Spec}k)\times\cdots\times(\mathbb{A}^1-0)) \to W$$

(where the ith factor has been replaced by $\operatorname{Spec}k$). The cofiber W/J should be homotopy equivalent to the motive $\mathbb{Z}(n)[n]$. Think about the analogous fact in topology: we are starting with the motive of a torus and quotienting out the motives of all smaller torii, and this should leave a single Eilenberg-MacLane space corresponding to the top piece of our original motive. For any scheme X, Voevodsky then writes down a chain complex serving as an algebraic model for a mapping space $\operatorname{Map}(M(X),W/J)$. How should the homology groups of this chain complex be related to motivic cohomology? We can guess the answer using what we've learned so far:

$$\begin{aligned}\pi_i \operatorname{Map}(M(X),W/J) &\cong \operatorname{Ho}(\mathcal{M}(k))(M(X)[i],Z(n)[n])\\ &\cong \operatorname{Ho}(\mathcal{M}(k))(M(X),Z(n)[n-i])\\ &\cong H^{n-i,n}(X).\end{aligned}$$

In another paper Voevodsky considers something like the quotient motive $Q = M(\mathbb{P}^n)/M(\mathbb{P}^{n-1})$ and write down an algebraic model for the mapping space $\operatorname{Map}(M(X),Q)$. The homology groups of this chain complex are again related to motivic cohomology, and we can figure out how. The motive of $\mathbb{P}^n$ splits as

$$M(\mathbb{P}^n) \simeq \mathbb{Z}(0)\oplus\mathbb{Z}(1)[2]\oplus\mathbb{Z}(2)[4]\oplus\cdots\oplus\mathbb{Z}(n)[2n].$$

The submotive $M(\mathbb{P}^{n-1})$ consists of all the factors except the last one, so that

$$Q \simeq \mathbb{Z}(n)[2n].$$

Then we should have

$$\begin{aligned}\pi_i \operatorname{Map}(M(X),Q) \cong \operatorname{Ho}(\mathcal{M}(k))(M(X)[i],\mathbb{Z}(n)[2n]) &\cong \operatorname{Ho}(\mathcal{M}(k))(M(X),\mathbb{Z}(n)[2n-i])\\ &\cong H^{2n-i,n}(X).\end{aligned}$$

In yet another context Voevodsky writes down an algebraic model for the mapping space $\operatorname{Map}(M(X),\widetilde{M}(\mathbb{A}^n-0))$. We know that we should have $\widetilde{M}(\mathbb{A}^n-0) \simeq \mathbb{Z}(n)[2n-1]$ and so

$$\begin{aligned}\pi_i \operatorname{Map}(M(X),\widetilde{M}(A^n-0)) &\cong \operatorname{Ho}(\mathcal{M}(k))(M(X)[i],\mathbb{Z}(n)[2n-1])\\ &\cong \operatorname{Ho}(\mathcal{M}(k))(M(X),\mathbb{Z}(n)[2n-i-1])\\ &\cong H^{2n-i-1,n}(X).\end{aligned}$$

CHAPTER 10

Crystalline cohomology

Nothing here yet.

CHAPTER 11

The Milnor conjectures

These lectures concern the two Milnor conjectures and their proofs: from [**V3**], [**OVV**], and [**M2**]. Voevodsky's proof of the norm residue symbol conjecture—which is now eight years old—came with an explosion of ideas. The aim of these notes is to make this explosion a little more accessible to topologists. My intention is not to give a completely rigorous treatment of this material, but just to outline the main ideas and point the reader in directions where he can learn more. I've tried to make the lectures accessible to topologists with no specialized knowledge in this area, at least to the extent that such a person can come away with a general sense of how homotopy theory enters into the picture.

Let me apologize for two aspects of these notes. Foremost, they reflect only my own limited understanding of this material. Secondly, I have made certain expository decisions about which parts of the proofs to present in detail and which parts to keep in a "black box"—and the reader may well be disappointed in my choices. I hope that in spite of these shortcomings the notes are still useful.

Sections 1, 2, and 3 each depend heavily on the previous one. Section 4 could almost be read independently of 2 and 3, except for the need of Remark 2.10.

1. The conjectures

The Milnor conjectures are two purely algebraic statements in the theory of fields, having to do with the classification of quadratic forms. In this section we'll review the basic theory and summarize the conjectures. Appendix A contains some supplementary material, where several examples are discussed.

1.1. Background. Let F be a field. In some sense our goal is to completely classify symmetric bilinear forms over F. To give such a form $(-,-)$ on F^n is the same as giving a symmetric $n \times n$ matrix A, where $a_{ij} = (e_i, e_j)$. Two matrices A_1 and A_2 represent the same form up to a change of basis if and only if $A_1 = PA_2P^T$ for some invertible matrix P. The main classical theorem on this topic says that if $\mathrm{char}(F) \neq 2$ then every symmetric bilinear form can be diagonalized by a change of basis. The question remains to decide when two given diagonal matrices D_1 and D_2 represent equivalent bilinear forms. For instance, do $\begin{bmatrix} 2 & 0 \\ 0 & 11 \end{bmatrix}$ and $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ represent the same form over $\mathbb{Q}$?

To pursue this question one looks for invariants. The most obvious of these is the rank of the matrix A. This is in fact the unique invariant when the field is algebraically closed. For suppose a form is represented by a diagonal matrix D, and let λ be a nonzero scalar. Construct a new basis by replacing the ith basis element e_i by λe_i. The matrix of the form with respect to this new basis is the same as D, but with the ith diagonal entry multiplied by λ^2. The conclusion is that multiplying the entries of D by squares does not change the isomorphism class of the underlying form. This leads immediately to the classical theorem saying that if every element of F is a square (which we'll write as $F = F^2$) then a symmetric bilinear form is completely classified by its rank.

We now restrict to nondegenerate forms, in which case the matrix A is nonsingular. The element $\det(A) \in F^*$ is not quite an invariant of the bilinear form, since after a change of basis the determinant of the new matrix will be $\det(P)\det(A)\det(P^T) = \det(P)^2\det(A)$. However, the determinant is a well-defined invariant if we regard it as an element of $F^*/(F^*)^2$. Since $\frac{22}{3}$ is not a square in $\mathbb{Q}$, for instance, this tells us that the matrices $\begin{bmatrix} 2 & 0 \\ 0 & 11 \end{bmatrix}$ and $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ don't represent isomorphic forms over $\mathbb{Q}$.

The rank and determinant are by far the simplest invariants to write down, but they are not very strong. They don't even suffice to distinguish forms over $\mathbb{R}$. This case is actually a good example to look at. For $a_1, \ldots, a_n \in \mathbb{R}^*$, let $\langle a_1, \ldots, a_n \rangle$ denote the form on $\mathbb{R}^n$ defined by $(e_i, e_j) = \delta_{i,j} a_i$. Since every element of $\mathbb{R}$ is either a square or the negative of a square, it follows that every nondegenerate real form is isomorphic to an $\langle a_1, \ldots, a_n \rangle$ where each $a_i \in \{1, -1\}$. When are two such forms isomorphic? Of course one knows the answer, but let's think through it. The Witt Cancellation Theorem (true over any field) says that if $\langle x_1, \ldots, x_n, y_1, \ldots, y_k \rangle \cong \langle x_1, \ldots, x_n, z_1, \ldots, z_k \rangle$ then $\langle y_1, \ldots, y_k \rangle \cong \langle z_1, \ldots, z_k \rangle$. So our problem reduces to deciding whether the n-dimensional forms $\langle 1, 1, \ldots, 1 \rangle$ and $\langle -1, \ldots, -1 \rangle$ are isomorphic. When n is odd the determinant distinguishes them, but when n is even it doesn't. Of course the thing to say is that the associated quadratic form takes only positive values in the first case, and only negative values in the second—but this is not exactly an 'algebraic' way of distinguishing the forms, in that it uses the ordering on $\mathbb{R}$ in an essential way. By the end of this section we will indeed have purely algebraic invariants we can use here.

1.2. The Grothendieck-Witt ring. In a moment we'll return to the problem of finding invariants more sophisticated than the rank and determinant, but first we need a little more machinery. From now on $\mathrm{char}(F) \neq 2$. By a **quadratic space** I mean a pair (V, μ) consisting of a finite-dimensional vector space and a non-degenerate bilinear form μ. To systemize their study one defines the *Grothendieck-Witt ring* $GW(F)$. This is the free abelian group generated by isomorphism classes of pairs (V, μ), with the usual relation identifying the direct sum of quadratic spaces with the sum in the group. The multiplication is given by tensor product of vector spaces.

The classical theory of bilinear forms allows us to give a complete description of the abelian group $GW(F)$ in terms of generators and relations. Recall that $\langle a_1, \ldots, a_n \rangle$ denotes the n-dimensional space F^n with $(e_i, e_j) = \delta_{ij} a_i$. So $\langle a_1, \ldots, a_n \rangle = \langle a_1 \rangle + \cdots + \langle a_n \rangle$ in $GW(F)$. The fact that every symmetric bilinear form is diagonalizable tells us that $GW(F)$ is generated by the elements $\langle a \rangle$ for $a \in F^*$, and we have already observed the relation $\langle ab^2 \rangle = \langle a \rangle$ for any $a, b \in F^*$. As an easy exercise, one can also give a complete description for when *two-dimensional* forms are isomorphic: one must be able to pass from one to the other via the two relations

$$\langle ab^2\rangle = \langle a\rangle \quad \text{and} \quad \langle a,b\rangle = \langle a+b, ab(a+b)\rangle \tag{1.3}$$

where in the second we assume $a, b \in F^*$ and $a+b \neq 0$. As an example, working over $\mathbb{Q}$ we have

$$\langle 3,-2\rangle = \langle 12,-2\rangle = \langle 10,-240\rangle = \langle 90,-15\rangle.$$

To completely determine all relations in $GW(F)$, one shows that if two forms $\langle a_1,\dots,a_n\rangle$ and $\langle b_1,\dots,b_n\rangle$ are isomorphic then there is a chain of isomorphic diagonal forms connecting one to the other, where each link of the chain differs in exactly two elements. Thus, (1.3) is a complete set of relations for $GW(F)$. The reader may consult [**S1**, 2.9.4] for complete details here.

The multiplication in $GW(F)$ can be described compactly by

$$\langle a_1,\dots,a_n\rangle \cdot \langle b_1,\dots,b_k\rangle = \sum_{i,j} \langle a_i b_j\rangle.$$

1.4. The Witt ring. The *Witt ring* $W(F)$ is the quotient of $GW(F)$ by the ideal generated by the so-called 'hyperbolic plane' $\langle 1,-1\rangle$. Historically $W(F)$ was studied long before $GW(F)$, probably because it can be defined without formally adjoining additive inverses as was done for $GW(F)$. One can check that the forms $\langle a,-a\rangle$ and $\langle 1,-1\rangle$ are isomorphic, and therefore if one regards hyperbolic forms as being zero then $\langle a_1,\dots,a_n\rangle$ and $\langle -a_1,\dots,-a_n\rangle$ are additive inverses. So $W(F)$ can be described as a set of equivalence classes of quadratic spaces, and doesn't require working with 'virtual' objects.

Because $\langle a,-a\rangle \cong \langle 1,-1\rangle$ for any a, it follows that the ideal $\big(\langle 1,-1\rangle\big)$ is precisely the additive subgroup of $GW(F)$ generated by $\langle 1,-1\rangle$. As an abelian group, it is just a copy of $\mathbb{Z}$. So we have the exact sequence $0 \to \mathbb{Z} \to GW(F) \to W(F) \to 0$.

Let $GI(F)$ be the kernel of the dimension function $\dim\colon GW(F) \to \mathbb{Z}$, usually called the augmentation ideal. Let $I(F)$ be the image of the composite $GI(F) \hookrightarrow GW(F) \twoheadrightarrow W(F)$; one can check that $I(F)$ consists precisely of equivalence classes of even-dimensional quadratic spaces. Note that I is additively generated by forms $\langle 1,a\rangle$, and therefore I^n is additively generated by n-fold products $\langle 1,a_1\rangle\langle 1,a_2\rangle\cdots\langle 1,a_n\rangle$.

The dimension function gives an isomorphism $W/I \to \mathbb{Z}/2$. The determinant gives us a group homomorphism $GW(F) \to F^*/(F^*)^2$, but it does not extend to the Witt ring because $\det\langle 1,-1\rangle = -1$. One defines the *discriminant* of $\langle a_1,\dots,a_n\rangle$ to be $(-1)^{\frac{n(n-1)}{2}} \cdot (a_1\cdots a_n)$, and with this definition the discriminant gives a map of sets $W(F) \to F^*/(F^*)^2$. It is not a homomorphism, but if we restrict to $I(F) \to F^*/(F^*)^2$ then it *is* a homomorphism. As the discriminant of $\langle 1,a\rangle\langle 1,b\rangle$ is a square, the elements of I^2 all map to 1. So we get an induced map $I/I^2 \to F^*/(F^*)^2$, which is obviously surjective. It is actually an isomorphism—to see this, note that

$$\langle x,y\rangle\langle -1,y\rangle = \langle -x,xy,-y,y^2\rangle = \langle 1,-x,-y,xy\rangle$$

and so $\langle x,y\rangle \equiv \langle 1,xy\rangle \pmod{I^2}$. It follows inductively that $\langle a_1,\dots,a_{2n}\rangle \equiv \langle 1,1,\dots,1,a_1a_2\cdots a_{2n}\rangle \pmod{I^2}$. So if $\langle a_1,\dots,a_{2n}\rangle$ is a form whose discriminant is a square, it is equivalent mod I^2 to either $\langle 1,1,\dots,1\rangle = 2n\langle 1\rangle$ (if n is even) or $\langle 1,1,\dots,1,-1\rangle = (2n-2)\langle 1\rangle$ (if n is odd). In the former case $2n\langle 1\rangle = 2\langle 1\rangle \cdot n\langle 1\rangle \in I^2$, and in the latter case $(2n-2)\langle 1\rangle = 2\langle 1\rangle \cdot (n-1)\langle 1\rangle \in I^2$. In either case we have $\langle a_1,\dots,a_{2n}\rangle \in I^2$, and this proves injectivity.

The examples in the previous paragraph are very special, but they suggest why one might hope for 'higher' invariants which give isomorphisms between the groups I^n/I^{n+1} and something more explicitly defined in terms of the field F. This is what the Milnor conjecture is about.

REMARK 1.5. For future reference, note that $2\langle 1\rangle = \langle 1,1\rangle \in I$, and therefore the groups I^n/I^{n+1} are $\mathbb{Z}/2$-vector spaces. Also observe that $GI(F)$ does not intersect the kernel of $GW(F) \to W(F)$, and so $GI(F) \to I(F)$ is an isomorphism. It follows that $(GI)^n/(GI)^{n+1} \cong I^n/I^{n+1}$, for all n.

1.6. More invariants. Recall that the **Brauer group** $\mathrm{Br}(F)$ is a set of equivalence classes of central, simple F-algebras, with the group structure coming from tensor product. The inverse of such an algebra is its opposite algebra, where the order of multiplication has been reversed.

From a quadratic space (V,μ) one can construct the associated Clifford algebra $C(\mu)$: this is the quotient of the tensor algebra $T_F(V)$ by the relations generated by $v\otimes v = \mu(v,v)$. Clifford algebras are $\mathbb{Z}/2$-graded by tensor length. If μ is even-dimensional then $C(\mu)$ is a central simple algebra, and if μ is odd-dimensional then the even part $C_0(q)$ is a central simple algebra. So we get an invariant of quadratic spaces taking its values in $\mathrm{Br}(F)$ (see [**S1**, 9.2.12] for more detail). This is usually called the **Clifford invariant**, or sometimes the **Witt invariant**. Since any Clifford algebra is isomorphic to its opposite, the invariant always produces a 2-torsion class.

Now we need to recall some Galois cohomology. Let $\bar{F}$ be a separable closure of F, and let $G = \mathrm{Gal}(\bar{F}/F)$. Consider the short exact sequence of G-modules $0 \to \mathbb{Z}/2 \to \bar{F}^* \to \bar{F}^* \to 0$, where the second map is squaring. Hilbert's Theorem 90 implies that $H^1(G;\bar{F}^*) = 0$, which means that the induced long exact sequence in Galois cohomology splits up into

$$0 \to H^0(G;\mathbb{Z}/2) \to F^* \xrightarrow{2} F^* \to H^1(G;\mathbb{Z}/2) \to 0$$

and

$$0 \to H^2(G;\mathbb{Z}/2) \to H^2(G;\bar{F}^*) \xrightarrow{2} H^2(G;\bar{F}^*).$$

The group $H^2(G;\bar{F}^*)$ is known to be isomorphic to $\mathrm{Br}(F)$, so we have $H^0(G;\mathbb{Z}/2) = \mathbb{Z}/2$, $H^1(G;\mathbb{Z}/2) = F^*/(F^*)^2$, and the 2-torsion in the Brauer group is precisely $H^2(G;\mathbb{Z}/2)$. From now on we will write $H^*(F;\mathbb{Z}/2) = H^*(G;\mathbb{Z}/2)$.

At this point we have the rank map $e_0\colon W(F) \to \mathbb{Z}/2 = H^0(F;\mathbb{Z}/2)$, which gives an isomorphism $W/I \to \mathbb{Z}/2$. We have the discriminant $e_1\colon I(F) \to F^*/(F^*)^2 = H^1(F;\mathbb{Z}/2)$ which gives an isomorphism $I/I^2 \to F^*/(F^*)^2$, and we have the Clifford invariant $e_2\colon I^2 \to H^2(F;\mathbb{Z}/2)$. With a little work one can check that e_2 is a homomorphism, and it kills I^3. The question of whether $I^2/I^3 \to H^2(F;\mathbb{Z}/2)$ is an isomorphism is difficult, and wasn't proven until the early 80s by Merkurjev [**M**] (neither surjectivity nor injectivity is obvious). The maps e_0, e_1, e_2 are usually called the *classical invariants* of quadratic forms.

The above isomorphisms can be rephrased as follows. The ideal I consists of all elements where $e_0 = 0$; I^2 consists of all elements such that $e_0 = 0$ and $e_1 = 1$; and by Merkujev's theorem I^3 is precisely the set of elements for which e_0, e_1, and e_2 are all trivial. Quadratic forms will be completely classified by these invariants if $I^3 = 0$, but unfortunately this is usually not the case. This brings us to the search for higher invariants. One early result along these lines is due to Delzant [**De**], who defined Stiefel-Whitney invariants with values in Galois cohomology. Unfortunately these are not the 'right' invariants, as they do not lead to complete classifications for elements in I^n, $n \geq 3$.

1.7. Milnor's work. At this point we find ourselves looking at the two rings $\operatorname{Gr}_I W(F)$ and $H^*(F;\mathbb{Z}/2)$, and we have maps between them in dimensions 0, 1, and 2. I think Milnor, inspired by his work on algebraic K-theory, wrote down the best ring he could find which would map to both rings above. In [**Mr2**] he defined what is now called 'Milnor K-theory' as

$$K_*^M(F) = T_{\mathbb{Z}}(F^*)/\langle a \otimes (1-a) | a \in F - \{0,1\}\rangle$$

where $T_{\mathbb{Z}}(V)$ denotes the tensor algebra over $\mathbb{Z}$ on the abelian group V. The grading comes from the grading on the tensor algebra, in terms of word length. I will write $\{a_1, \ldots, a_n\}$ for the element $a_1 \otimes \cdots \otimes a_n \in K_n^M(F)$.

Note that when dealing with $K_*^M(F)$ one must be careful not to confuse the addition—which comes from multiplication in F^*—with the multiplication. So for instance $\{a\}+\{b\} = \{ab\}$ but $\{a\}\cdot\{b\} = \{a,b\}$. This is in contrast to the operations in $GW(F)$, where one has $\langle a\rangle + \langle b\rangle = \langle a,b\rangle$ and $\langle a\rangle \otimes \langle b\rangle = \langle ab\rangle$. Unfortunately it is very easy to get these confused. Note that $\{a^2\} = 2\{a\}$, and more generally $\{a^2, b_1, \ldots, b_n\} = 2\{a, b_1, \ldots, b_n\}$.

Remark 1.8. From a modern perspective the name 'K-theory' applied to $K_*^M(F)$ is somewhat of a misnomer; one should not take it too seriously. The construction turns out to be more closely tied to algebraic cycles than to algebraic K-theory, and so I personally like the term 'Milnor cycle groups'. I doubt this will ever catch on, however.

Milnor produced two ring homomorphisms $\eta\colon K_*^M(F)/2 \to H^*(F;\mathbb{Z}/2)$ and $\nu\colon K_*^M(F)/2 \to \operatorname{Gr}_I W(F)$. To define the map ν, note first that we have already established an isomorphism $F^*/(F^*)^2 \to I/I^2$ sending $\{a\}$ to $\langle a,-1\rangle = \langle a\rangle - \langle 1\rangle$ (this is the inverse of the discriminant). This tells us what ν does to elements in degree 1. Since these elements generate $K_*^M(F)$ multiplicatively, to construct ν it suffices to verify that the appropriate relations are satisfied in the image. So we first need to check that

$$0 = \Big(\langle a\rangle - \langle 1\rangle\Big)\cdot\Big(\langle 1-a\rangle - \langle 1\rangle\Big) = \langle a(1-a)\rangle - \langle a\rangle - \langle 1-a\rangle + \langle 1\rangle = \langle a(1-a), 1\rangle - \langle a, 1-a\rangle,$$

but this follows directly from the second relation in (1.3). We also must check that $2\{a\}$ maps to 0, but $2\{a\} = \{a^2\} \mapsto \langle a^2\rangle - \langle 1\rangle$ and the latter vanishes by the first relation in (1.3). For future reference, note that $\nu(\{a\})$ is equal to both $\langle a,-1\rangle$ and $\langle -a, 1\rangle$ in I/I^2, since this group is 2-torsion.

Defining η is similar. We have already noticed that there is a natural isomorphism $H^1(F;\mathbb{Z}/2) \cong F^*/(F^*)^2$, and so it is clear where the element $\{a\}$ in $K_1^M(F) = F^*$ must be sent. The verification that $a \cup (1-a) = 0$ in $H^2(F;\mathbb{Z}/2)$ is in [**Mr2**, 6.1].

Milnor observed that both η and ν were isomorphisms in all the cases he could compute. The claim that η is an isomorphism is nowadays known as *the Milnor conjecture*, and was proven by Voevodsky in 1996 [**V1**]. The claim that ν is an isomorphism goes under the name *Milnor's conjecture on quadratic forms*. For characteristic 0 it was proven in 1996 by Orlov, Vishik, and Voevodsky [**OVV**], who deduced it as a consequence of the work in [**V1**]. I believe the proof now works in characteristic p, based on the improved results of [**V3**]. A second proof, also in characteristic 0, was outlined by Morel [**M2**] using the motivic Adams spectral sequence, and again depended on results from [**V1**]; unfortunately complete details of Morel's proof have yet to appear.

It is interesting that the conjecture on quadratic forms doesn't have an independent proof, and is the less primary of the two. Note that both $K_*^M(F)/2$ and $GW(F)$ can be completely described in terms of generators and relations (although the latter does not quite imply that we know all the relations in $\operatorname{Gr}_I W(F)$, which is largely the problem). The map ν is easily seen to be surjective, and so the only question is injectivity. Given this, it is in some ways surprising that the conjecture is as hard as it is.

REMARK 1.9. The map η is called the *norm residue symbol*, and can be defined for primes other than 2. The *Bloch-Kato conjecture* is the statement that $\eta\colon K_i^M(F)/l \to H^i(F;\mu_l^{\otimes i})$ is an isomorphism for l a prime different from char(F). This is a direct generalization of the Milnor conjecture to the case of odd primes. A proof was released by Voevodsky in 2003 [**V4**] (although certain auxiliary results required for the proof remain unwritten). I'm not sure anyone has ever considered an odd-primary analog of Milnor's conjecture on quadratic forms—what could replace the Grothendieck-Witt ring here?

At this point it might be useful to think through the Milnor conjectures in a few concrete examples. For these we refer the reader to Appendix A. Let's at least note here that through the work of Milnor, Bass, and Tate (cf. [**Mr2**]) the conjectures could be verified for all finite fields and for all finite extensions of $\mathbb{Q}$ (in fact for all global fields).

Finally, let's briefly return to the classification of forms over $\mathbb{R}$. We saw earlier that this reduces to proving that the n-dimensional forms $\langle 1,1,\ldots,1\rangle$ and $\langle -1,-1,\ldots,-1\rangle$ are not isomorphic. Can we now do this algebraically? If they were isomorphic, they would represent the same element of $W(\mathbb{R})$. It would follow that $(2n)\langle 1\rangle = 0$ in $W(\mathbb{R})$. Can this happen? The isomorphisms $\mathbb{Z}/2[a] \cong H^*(\mathbb{Z}/2;\mathbb{Z}/2) \cong K_*^M(\mathbb{R})/2 \cong \operatorname{Gr}_I W(\mathbb{R})$ show that $\operatorname{Gr}_I W(\mathbb{R})$ is a polynomial algebra on the class $\langle -1,-1\rangle$ (the generator a corresponds to the generator -1 of $\mathbb{R}^*/(\mathbb{R}^*)^2$, and $\nu(-1) = \langle -1,-1\rangle$). It follows that $2^k\langle 1\rangle = \pm\langle -1,-1\rangle^k$ is a generator for the group $I^k/I^{k+1} \cong \mathbb{Z}/2$. If $m = 2^i r$ where r is odd, then $m\langle 1\rangle = 2^i\langle 1\rangle \cdot r\langle 1\rangle$. Since $r\langle 1\rangle$ is the generator for W/I and $2^i\langle 1\rangle$ is a generator for I^i/I^{i+1}, it follows that $m\langle 1\rangle$ is also a generator for I^i/I^{i+1}. In particular, $m\langle 1\rangle$ is nonzero. So we have proven via algebraic methods (although in this case also somewhat pathological ones) that $\langle 1,1,\ldots,1\rangle \ncong \langle -1,-1,\ldots,-1\rangle$.

1.10. Further background reading. There are several good expository papers on the theory of quadratic forms, for example [**Pf1**] and [**S2**]. The book [**S1**] is a very thorough and readable resource as well. For the Milnor conjectures themselves there is [**Pf2**], which in particular gives several applications of the conjectures; it also gives detailed references to original papers. The beginning sections of [**AEJ**] offer a nice survey concerning the search for 'higher' invariants of quadratic forms. It's worth pointing out that after Milnor's work definitions of e_3, e_4, and e_5 were eventually given—with a lot of hard work—but this was the state of the art until 1996. Finally, the introduction of [**V3**] gives a history of work on the Milnor conjecture.

2. Proof of the conjecture on the norm residue symbol

This section outlines Voevodsky's proof of the Milnor conjecture on the norm residue symbol [**V1, V3**]. Detailed, step-by-step summaries have been given in [**M1**] and [**Su**]. My intention here is not to give a complete, mathematically rigorous presentation, but rather just to give the flavor of what is involved.

Several steps in the proof involve manipulations with motivic cohomology based on techniques that were developed in [**VSF**]. I have avoided giving any details about these steps, in an attempt to help the exposition. Most of these details are not hard to understand, however—there are only a few basic techniques to keep track of, and one can read about them in [**VSF**] or [**MVW**]. But I hope that by keeping some of this stuff in a black box the overall structure of the argument will become clearer.

2.1. Initial observations. The aim is to show that $\eta\colon K_*^M(F)/2 \to H^*(F;\mathbb{Z}/2)$ is an isomorphism. To do this, one of the first things one might try to figure out is what kind of extra structure $K_*^M(F)/2$ and $H^*(F;\mathbb{Z}/2)$ have in common. For instance, they are both covariant functors in F, and the covariance is compatible with the norm residue symbol. It turns out they both have transfer maps for finite separable extensions (which, for those who like to think geometrically, are the analogs of covering spaces). That is, if $j\colon F \hookrightarrow F'$ is a separable extension of degree n then there is a map $j^!\colon K_*^M(F') \to K_*^M(F)$ such that $j^! j_*$ is multiplication by n, and similarly for $H^*(F;\mathbb{Z}/2)$. (Note that the construction of transfer maps for Milnor K-theory is not at all trivial—some ideas were given in [**BT**, Sec. 5.9], but the full construction is due to Kato [**K1**, Sec. 1.7]). It follows that if n is odd then $K_*^M(F)/2 \to H^*(F;\mathbb{Z}/2)$ is a retract of the map $K_*^M(F')/2 \to H^*(F';\mathbb{Z}/2)$. So if one had a counterexample to the Milnor conjecture, field extensions of all odd degrees would still be counterexamples. This is often referred to as "the transfer argument".

Another observation is that both functors can be extended to rings other than fields, and if R is a discrete valuation ring then both functors have a 'localization sequence' relating their values on R, the residue field, and the quotient field. I will not go into details here, but if F is a field of characteristic p then by using the Witt vectors over F and the corresponding localization sequence, one can reduce the Milnor conjecture to the case of characteristic 0 fields. The argument is in [**V1**, Lemma 5.2]. In Voevodsky's updated proof of the Milnor conjecture [**V3**] this step is not necessary, but I think it's useful to realize that the Milnor conjecture is not hard because of 'crazy' things that might happen in characteristic p—it is hard even in characteristic 0.

2.2. A first look at the proof. The proof goes by induction. We assume the norm residue map $\eta\colon K_*^M(F)/2 \to H^*(F;\mathbb{Z}/2)$ is an isomorphism for all fields F and all $* < n$, and then prove it is also an isomorphism for $* = n$. The basic theme of the proof, which goes back to Merkurjev, involves two steps:

(1) Verify that η_n is an isomorphism for certain 'big' fields—in our case, those which have no extensions of odd degree and also satisfy $K_n(F) = 2K_n(F)$ (so that one must prove $H^n(F;\mathbb{Z}/2) = 0$). Notice that when $n = 1$ the condition $K_1 = 2K_1$ says that $F = F^2$.
(2) Prove that if F were a field for which η_n is *not* an isomorphism then one could expand F to make a 'bigger' counterexample, and could keep doing this until

you're in the range covered by step (1). This would show that no such F could exist.

In more detail one shows that for any $\{a_1, \dots, a_n\} \in K_n(F)$ one can construct an extension $F \hookrightarrow F'$ with the property that $\{a_1, \dots, a_n\} \in 2K_n(F')$ and $\eta_n \colon K_n(F')/2 \to H^n(F'; \mathbb{Z}/2)$ still fails to be an isomorphism. By doing this over and over and taking a big colimit, one gets a counterexample where $K_n^M = 2K_n^M$.

Neither of the above two steps is trivial, but step (1) involves nothing very fancy—it is a calculation in Galois cohomology which takes a few pages, but is not especially hard. See [**V3**, Section 5]. Step (2) is the more subtle and interesting step. Note that if $\underline{a} = \{a_1, \dots, a_n\} \notin 2K_n^M(F)$ then none of the a_i's can be in F^2. There are several ways one can extend F to a field F' such that $\underline{a} \in 2K_n^M(F')$: one can adjoin a square root of any a_i, for instance. The problem is to find such an extension where you have enough control over the horizontal maps in the diagram

$$\begin{tikzcd} K_n^M(F)/2 \arrow[r] \arrow[d, "\eta_F"'] & K_n^M(F')/2 \arrow[d, "\eta_{F'}"] \\ H^n(F; \mathbb{Z}/2) \arrow[r] & H^n(F'; \mathbb{Z}/2) \end{tikzcd}$$

to show that if η_F fails to be an isomorphism then so does $\eta_{F'}$. The selection of the 'right' F' is delicate.

We will alter our language at this point, because we will want to bring more geometry into the picture. Any finitely-generated separable extension $F \hookrightarrow F'$ is the function field of a smooth F-variety. A **splitting variety** for an element $\underline{a} \in K_n^M(F)$ is a smooth variety X, of finite type over F, with the property that $\underline{a} \in 2K_n^M(F(X))$. Here $F(X)$ denotes the function field of X. As we just remarked, there are many such varieties: $X = \operatorname{Spec} F[u]/(u^2 - a_1)$ is an example. The particular choice we'll be interested in is more complicated.

Given $b_1, \dots, b_k \in F$, let $q_{\underline{b}}$ be the quadratic form in 2^k variables corresponding to the element

$$\langle 1, -b_1 \rangle \otimes \langle 1, -b_2 \rangle \otimes \cdots \otimes \langle 1, -b_k \rangle \in GW(F).$$

For example, $q_{b_1,b_2}(x_1, \dots, x_4) = x_1^2 - b_1x_2^2 - b_2x_3^2 + b_1b_2x_4^2$. Such q's are called **Pfister forms**, and they have a central role in the modern theory of quadratic forms (see [**S1**, Chapter 4], for instance).

For $a_1, \dots, a_n \in F$, define $Q_{\underline{a}}$ to be the projective quadric in $\mathbb{P}^{2^{n-1}}$ given by the equation

$$q_{a_1,\dots,a_{n-1}}(x_0, \dots, x_{[2^{n-1}-1]}) - a_n x_{2^{n-1}}^2 = 0.$$

In [**V3**] these are called **norm quadrics**. A routine argument [**V3**, Prop. 4.1] shows that $Q_{\underline{a}}$ is a splitting variety for $\underline{a}$. The reason for choosing to study this particular splitting variety will not be clear until later; isolating this object is one of the key aspects of the proof.

The name of the game will be to understand enough about the difference between $K_n^M(F)/2$ and $K_n^M(F(Q_{\underline{a}}))/2$ (as well as the corresponding Galois cohomology groups) to show that $K_n^M(F(Q_{\underline{a}}))/2 \to H^n(F(Q_{\underline{a}}); \mathbb{Z}/2)$ still fails to be an isomorphism. Voevodsky's argument uses motivic cohomology—of the quadrics $Q_{\underline{a}}$ and other objects—to 'bridge the gap' between $K_n^M(F)/2$ and $K_n^M(F(Q_{\underline{a}}))/2$.

2.3. Motivic cohomology enters the picture. Motivic cohomology is a bi-graded functor $X \mapsto H^{p,q}(X;\mathbb{Z})$ defined on the category of smooth F-schemes. Actually it is defined for all simplicial smooth schemes, as well as for more general objects. One of the lessons of the last ten years is that one can set up a model category which contains all these objects, and then a homotopy theorist can deal with them in much the same ways he deals with ordinary topological spaces. From now on I will do this implicitly (without ever referring to the machinery involved).

The coefficient groups $H^{p,q}(\operatorname{Spec} F;\mathbb{Z})$ vanish for $q<0$ and for $p>q\geq 0$. For us an important point is that the groups $H^{n,n}(\operatorname{Spec} F;\mathbb{Z})$ are canonically isomorphic to $K_n^M(F)$. Proving this is not simple! An account is given in [**MVW**, Lecture 5]. Finally, we note that one can talk about motivic cohomology with finite coefficients $H^{p,q}(X;\mathbb{Z}/n)$, related to integral cohomology via the exact sequence

$$\cdots \to H^{p,q}(X;\mathbb{Z}) \xrightarrow{\times n} H^{p,q}(X;\mathbb{Z}) \to H^{p,q}(X;\mathbb{Z}/n) \to H^{p+1,q}(X;\mathbb{Z}) \to \cdots$$

The sequence shows $H^{n,n}(\operatorname{Spec} F;\mathbb{Z}/2) \cong K_n^M(F)/2$ and $H^{p,q}(\operatorname{Spec} F;\mathbb{Z}/2)=0$ for $p>q\geq 0$.

Now, there is also an analagous theory $H_L^{p,q}(X;\mathbb{Z})$ which is called **Lichtenbaum (or étale) motivic cohomology**. There is a natural transformation $H^{p,q}(X;\mathbb{Z}) \to H_L^{p,q}(X;\mathbb{Z})$. The theory $H_L^{*,*}$ is the closest theory to $H^{*,*}$ which satisfies descent for the étale topology (essentially meaning that when $E\to B$ is an étale map there is a spectral sequence starting with $H_L^{*,*}(E)$ and converging to $H_L^{*,*}(B)$). The relation between $H^{*,*}$ and $H_L^{*,*}$ is formally analagous to that between a cohomology theory and a certain Bousfield localization of it. It is known that $H_L^{p,q}(X;\mathbb{Z}/n)$ is canonically isomorphic to étale cohomology $H_{et}^p(X;\mu_n^{\otimes q})$, if n is prime to char(F). From this it follows that $H_L^{p,q}(\operatorname{Spec} F;\mathbb{Z}/2)$ is the Galois cohomology group $H^p(F;\mathbb{Z}/2)$, for all q. At this point we can re-phrase the Milnor conjecture as the statement that the maps $H^{p,p}(\operatorname{Spec} F;\mathbb{Z}/2) \to H_L^{p,p}(\operatorname{Spec} F;\mathbb{Z}/2)$ are isomorphisms.

There are other conjectures about the relation between $H^{*,*}$ and $H_L^{*,*}$ as well. A conjecture of Lichtenbaum says that $H^{p,q}(X;\mathbb{Z}) \to H_L^{p,q}(X;\mathbb{Z})$ should be an isomorphism whenever $p\leq q+1$. Note that this would imply a corresponding statement for $\mathbb{Z}/n$-coefficients, and in particular would imply the Milnor conjecture. Also, since one knows $H^{n+1,n}(\operatorname{Spec} F;\mathbb{Z})=0$ Lichtenbaum's conjecture would imply that $H_L^{n+1,n}(\operatorname{Spec} F;\mathbb{Z})$ also vanishes. This latter statement was conjectured independently by both Beilinson and Lichtenbaum, and is known as a the **Generalized Hilbert's Theorem 90** (the case $n=1$ is a translation of the statement that $H^1_{Gal}(F;\bar{F}^*)=0$, which follows from the classical Hilbert's Theorem 90).

By knowing enough about how to work with motivic cohomology, Voevodsky was able to prove the following relation among these conjectures (as well as other relations which we won't need):

Proposition 2.4. *Fix an $n\geq 0$. Assume that $H_L^{k+1,k}(\operatorname{Spec} F;\mathbb{Z}_{(2)})=0$ for all fields F and all $0\leq k\leq n$. Then for any smooth simplicial scheme X over a field F, the maps $H^{p,q}(X;\mathbb{Z}/2) \to H_L^{p,q}(X;\mathbb{Z}/2)$ are isomorphisms when $q\geq 0$ and $p\leq q\leq n$; and they are monomorphisms for $p-1=q\leq n$. In particular, applying this when $p=q$ and $X=\operatorname{Spec} F$ verifies the Milnor conjecture in dimensions $\leq n$.*

It's worth pointing out that the proof uses nothing special about the prime 2, and so the statement is valid for all other primes as well.

For us, the importance of the above proposition is two-fold. First, it says that to prove the Milnor conjecture one only has to worry about the vanishing of one set of groups (the $H_L^{n+1,n}$'s) rather than two sets (the kernel and cokernel of η). Secondly, inductively assuming that the Generalized Hilbert's Theorem 90 holds up through dimension n is going to give us a lot more to work with than inductively assuming the Milnor conjecture up through dimension n. Instead of just knowing stuff about $H^{n,n}$ of fields, we know stuff about $H^{p,q}$ of any smooth simplicial scheme. The need for this extra information is a key feature of the proof.

2.5. Čech complexes. We only need one more piece of machinery before returning to the proof of the Milnor conjecture. This piece is hard to motivate, and its introduction is one of the more ingenious aspects of the proof. The reader will just have to wait and see how it arises in section 2.6 (see also Remark 3.10).

Let X be any scheme. The **Čech complex $\check{C}X$** is the simplicial scheme with $(\check{C}X)_n = X \times X \times \cdots \times X$ ($n+1$ factors) and the obvious face and degeneracies. This simplicial scheme can be regarded as augmented by the map $X \to \operatorname{Spec} F$.

For a topological space the realization of the associated Čech complex is always contractible—in fact, choosing any point of X allows one to write down a contracting homotopy for the simplicial space $\check{C}X$. But in algebraic geometry the scheme X may not have rational points; i.e., there may not exist any maps $\operatorname{Spec} F \to X$ at all! If X does have a rational point then the same trick lets one write down a contracting homotopy, and therefore $\check{C}X$ behaves as if it were $\operatorname{Spec} F$ in all computations. (More formally, $\check{C}X$ is homotopy equivalent to $\operatorname{Spec} F$ in the motivic homotopy category).

Working in the motivic homotopy category, one finds that for any smooth scheme Y the set of homotopy classes $[Y, \check{C}X]$ is either empty or a singleton. The latter holds precisely if Y admits a Zariski cover $\{U_\alpha\}$ such that there exist scheme maps $U_\alpha \to X$ (not necessarily compatible on the intersections). The object $\check{C}X$ has no 'higher homotopy information', only this very simple discrete information about whether or not certain maps exist. One should think of $\check{C}X$ as very close to being contractible. I point out again that in topology there is always at least one map between nonempty spaces, and so $\check{C}X$ is not very interesting.

If $E \to B$ is an étale cover, then there is a spectral sequence whose input is $H_L^{*,*}(E;\mathbb{Z})$ and which converges to $H_L^{*,*}(B;\mathbb{Z})$ (this is the étale descent property). In particular, if X is a smooth scheme and we let $F' = F(X)$, $X' = X \times_F F'$, then $X' \to X$ and $\operatorname{Spec} F' \to \operatorname{Spec} F$ are both étale covers. The scheme X' necessarily has a rational point over F', so $\check{C}X'$ and $\operatorname{Spec} F'$ look the same to H_L. The étale descent property then shows that $\check{C}X$ and $\operatorname{Spec} F$ also look the same: in other words, the maps $H_L^{p,q}(\operatorname{Spec} F;\mathbb{Z}) \to H_L^{p,q}(\check{C}X;\mathbb{Z})$ are all isomorphisms (and the same for finite coefficients). This is *not* true for $H^{*,*}$ in place of $H_L^{*,*}$. One might paraphrase all this by saying that in the étale world $\check{C}X$ is contractible, just as it is in topology.

2.6. The proof. Now I am going to give a complete summary of the proof as it appears in [**V1, V3**]. Instead of proving the Milnor conjecture in its original form one instead concentrates on the more manageable conjecture that $H_L^{i+1,i}(\operatorname{Spec} F;\mathbb{Z}_{(2)}) = 0$ for all i and all fields F. One assumes this has been proven in the range $0 \le i < n$, and then shows that it also follows for $i = n$.

Suppose that F is a field with $H_L^{n+1,n}(F;\mathbb{Z}_{(2)}) \neq 0$. The transfer argument shows that any extension field of odd degree would still be a counterexample, so we can assume F has no extensions of odd degree. One checks via some Galois cohomology computations—see [**V3**, section 5]—that if such a field has $K_n^M(F) = 2K_n^M(F)$ then $H_L^{n+1,n}(\operatorname{Spec} F;\mathbb{Z}_{(2)}) = 0$. So our counterexample cannot have $K_n^M(F) = 2K_n^M(F)$. By the reasoning from section 2.2, it will suffice to show that for every $a_1, \ldots, a_n \in F$ the field $F(Q_{\underline{a}})$ is still a counterexample. We will in fact show that $H_L^{n+1,n}(F;\mathbb{Z}_{(2)}) \to H_L^{n+1,n}(F(Q_{\underline{a}});\mathbb{Z}_{(2)})$ is injective.

Suppose u is in the kernel of the above map, and consider the diagram

$$\begin{tikzcd}
 & H_L^{n+1,n}(\operatorname{Spec} F;\mathbb{Z}_{(2)}) \arrow[r] \arrow[d, "\cong"] & H_L^{n+1,n}(\operatorname{Spec} F(Q_{\underline{a}});\mathbb{Z}_{(2)}) \\
H^{n+1,n}(\check{C}Q_{\underline{a}};\mathbb{Z}_{(2)}) \arrow[r] & H_L^{n+1,n}(\check{C}Q_{\underline{a}};\mathbb{Z}_{(2)}). &
\end{tikzcd}$$

Let u' denote the image of u in $H_L^{n+1,n}(\check{C}Q_{\underline{a}};\mathbb{Z}_{(2)})$. One can show (after some extensive manipulations with motivic cohomology) that the hypothesis on u implies that u' is the image of an element in $H^{n+1,n}(\check{C}Q_{\underline{a}};\mathbb{Z}_{(2)})$. It will therefore be sufficient to show that this group is zero.

Let $\tilde{C}$ be defined by the cofiber sequence $(\check{C}Q_{\underline{a}})_+ \to (\operatorname{Spec} F)_+ \to \tilde{C}$. This means $\tilde{H}^{*,*}(\tilde{C})$ fits in an exact sequence

$$\to H^{p-1,q}(\check{C}Q_{\underline{a}}) \to \tilde{H}^{p,q}(\tilde{C}) \to H^{p,q}(\operatorname{Spec} F) \to H^{p,q}(\check{C}Q_{\underline{a}}) \to \tilde{H}^{p+1,q}(\tilde{C}) \to \cdots$$

So the reduced motivic cohomology of $\tilde{C}$ detects the 'difference' between the motivic cohomology of $\check{C}Q_{\underline{a}}$ and $\operatorname{Spec} F$. The fact that $H^{i,n}(\operatorname{Spec} F;\mathbb{Z}) = 0$ for $i > n$ shows that $H^{n+1,n}(\check{C}Q_{\underline{a}};\mathbb{Z}_{(2)}) \cong \tilde{H}^{n+2,n}(\tilde{C};\mathbb{Z}_{(2)})$. Since $Q_{\underline{a}}$ has a rational point (and therefore $\check{C}Q_{\underline{a}}$ is contractible) over a degree 2 extension of F, it follows from the transfer argument that the above group is killed by 2. To show that the group is zero it is therefore sufficient to prove that the image of $\tilde{H}^{n+2,n}(\tilde{C};\mathbb{Z}_{(2)}) \to \tilde{H}^{n+2,n}(\tilde{C};\mathbb{Z}/2)$ is zero. This is the same as the image of $\tilde{H}^{n+2,n}(\tilde{C};\mathbb{Z}) \to \tilde{H}^{n+2,n}(\tilde{C};\mathbb{Z}/2)$, which I'll denote by $\tilde{H}_{int}^{n+2,n}(\tilde{C};\mathbb{Z}/2)$.

So far most of what we have done is formal; but now we come to the crux of the argument. For any smooth scheme X one has cohomology operations acting on $H^{*,*}(X;\mathbb{Z}/2)$. In particular, one can produce analogs of the Steenrod operations: the Bockstein acts with bi-degree $(1,0)$, and Sq^{2^i} acts with bi-degree $(2^i, 2^{i-1})$. From these one defines the Milnor Q_i's, which have bi-degree $(2^{i+1}-1, 2^i-1)$. In ordinary topology these are defined inductively by $Q_0 = \beta$ and $Q_i = [Q_{i-1}, Sq^{2^i}]$, whereas motivically one has to add some extra terms to this commutator (these arise because the motivic cohomology of a point is nontrivial). One shows that $Q_i \circ Q_i = 0$, and that $Q_i = \beta q + q\beta$ for a certain operation q. It follows from the latter formula that Q_i maps elements in $\tilde{H}_{int}$ to elements in $\tilde{H}_{int}$. All of these facts also work in ordinary topology, it's just that the proofs here are a little more complex.

The next result is [**V3**, Cor. 3.8]. It is the first of two main ingredients needed to complete the proof.

PROPOSITION 2.7. *Let X be a smooth quadric in $\mathbb{P}^{2^n}$, and let $\tilde{C}X$ be defined by the cofiber sequence $(\check{C}X)_+ \to (\operatorname{Spec} F)_+ \to \tilde{C}X$. Then for $i \leq n$, every element of $\tilde{H}^{*,*}(\tilde{C}X; \mathbb{Z}/2)$ that is killed by Q_i is also in the image of Q_i.*

This is a purely 'topological' result, in that its proof uses no algebraic geometry. It follows from the most basic properties of the Steenrod operations, motivic cohomology (like Thom isomorphism), and elementary facts about the characteristic numbers of quadrics. The argument is purely homotopy-theoretic.

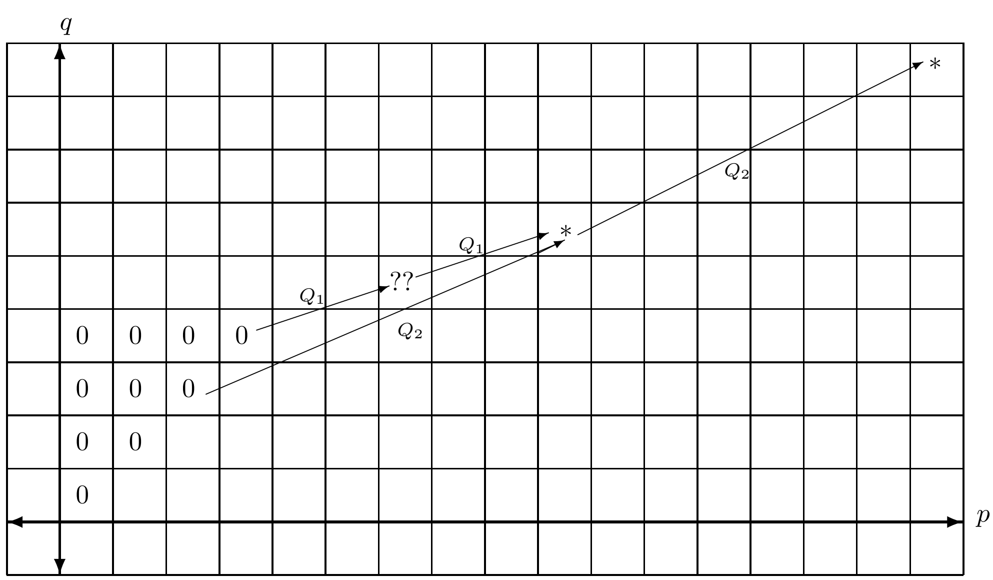
The second main result we will need is where all the algebraic geometry enters the picture. Voevodsky deduces it from results of Rost, who showed that the motive of $Q_{\underline{a}}$ splits off a certain direct summand. See [**V3**, Th. 4.9].

PROPOSITION 2.8. $\tilde{H}^{2^n,2^{n-1}}(\tilde{C}; \mathbb{Z}_{(2)}) = 0$.

Using the above two propositions we can complete the proof of the Milnor conjecture. In order to draw a concrete picture, let us just assume $n = 4$ for the moment. We are trying to show that $\tilde{H}^{6,4}_{int}(\tilde{C}; \mathbb{Z}/2) = 0$. Consider the diagram

$$\begin{tikzcd}
H^{p,q}(\operatorname{Spec} F; \mathbb{Z}/2) \arrow[r] \arrow[d] & H^{p,q}(\check{C}Q_{\underline{a}}; \mathbb{Z}/2) \arrow[d] \\
H^{p,q}_L(\operatorname{Spec} F; \mathbb{Z}/2) \arrow[r, "\cong"] & H^{p,q}_L(\check{C}Q_{\underline{a}}; \mathbb{Z}/2).
\end{tikzcd}$$

Our inductive assumption together with Proposition 2.4 implies that the vertical maps are isomorphisms for $p \leq q \leq n-1$, and monomorphisms for $p-1 = q \leq n-1$. So the top horizontal map is an isomorphism in the first range and a monomorphism in the second. The long exact sequence in motivic cohomology then shows that $\tilde{H}^{p,q}(\tilde{C}; \mathbb{Z}/2) = 0$ for $p \leq q \leq n-1$. This is where our induction hypothesis has gotten us. The following diagram depicts what we now know about $\tilde{H}^{p,q}(\tilde{C}; \mathbb{Z}/2)$ (the group marked ?? is $\tilde{H}^{6,4}$, the one we care about):



At this point Proposition 2.7 shows that $Q_1\colon H^{6,4} \to H^{9,5}$ is injective, and that $Q_2\colon H^{9,5} \to H^{16,8}$ is injective. Since the Q_i's take integral elements to integral elements, we have an inclusion

$$Q_2Q_1\colon \tilde{H}^{6,4}_{int}(\tilde{C};\mathbb{Z}/2) \hookrightarrow \tilde{H}^{16,8}_{int}(\tilde{C};\mathbb{Z}/2).$$

But it follows directly from Proposition 2.8 that $\tilde{H}^{16,8}_{int}(\tilde{C};\mathbb{Z}/2) = 0$, and so we are done.

The argument for general n follows exactly this pattern: one uses the composite of the operations $Q_1, Q_2, \ldots, Q_{n-2}$, but everything else is the same.

2.9. Summary. Here is a list of some of the key elements of the proof:

(1) The re-interpretation of the Milnor conjecture as a comparison of different bi-graded motivic cohomology theories. An extensive knowledge about such theories allows one to deduce statements for any smooth simplicial scheme from statements only about fields (cf. Proposition 2.4).
(2) Choice of the splitting variety $Q_{\underline{a}}$ (needed for Propositions 2.7 and 2.8).
(3) The introduction and use of Čech complexes.
(4) The construction of Steenrod operations on motivic cohomology and development of their basic properties, leading to the proof of Proposition 2.7.
(5) The 'geometric' results of Rost on motives of quadrics, which lead to Proposition 2.8.

2.10. A notable consequence. The integral motivic cohomology groups of a point $H^{p,q}(\operatorname{Spec} F)$ are largely unknown—the exception is when $q = 0, 1$. However, the proof of the Milnor conjecture tells us exactly what $H^{p,q}(\operatorname{Spec} F;\mathbb{Z}/2)$ is. First of all, independently of the Milnor conjecture it can be shown to vanish when $p \geq q$ and when $q < 0$. By Proposition 2.4 (noting that we now know the hypothesis to be satisfied for all n), it follows that

$$H^{p,q}(\operatorname{Spec} F;\mathbb{Z}/2) \to H^p_{et}(\operatorname{Spec} F;\mu_2^{\otimes q})$$

is an isomorphism when $p \leq q$ and $q \geq 0$. As $\mu_2^{\otimes q} \cong \mu_2$, the étale cohomology groups are periodic in q; that is, $H^*_{et}(\operatorname{Spec} F;\mu_2^{\otimes *}) \cong H^*_{Gal}(F;\mathbb{Z}/2)[\tau,\tau^{-1}]$ where τ has degree $(0,1)$.

The conclusion is that $H^{*,*}(\operatorname{Spec} F;\mathbb{Z}/2) \cong H^*_{Gal}(F;\mathbb{Z}/2)[\tau]$, where τ is the canonical class in $H^{0,1}$ and the Galois cohomology is regarded as the subalgebra lying in degrees (k,k). Of course the Milnor conjecture tells us that the Galois cohomology is the same as mod 2 Milnor K-theory, and so we can also write $H^{*,*}(\operatorname{Spec} F;\mathbb{Z}/2) \cong \big(K^M_*(F)/2\big)[\tau]$.

2.11. Further reading. Both the original papers of Voevodsky [**V1, V3**] are very readable, and remain the best sources for the proof. Summaries have also been given in [**M1**] and [**Su**]. A proof of the general Bloch-Kato conjecture was recently given in [**V4**]—the proof is similar in broad outline to the 2-primary case we described here, but with several important differences. See the introduction to [**V4**].

Of course in this section I have completely avoided discussing the two main elements of the proof, namely Propositions 2.7 and 2.8. The proof of Proposition 2.7 is in [**V1, V3**] and is written in a way that can be understood by most homotopy theorists. Proposition 2.8 depends on results of Rost, which seem to be largely unpublished. See [**R1, R2**] for summaries.

For more about why Čech complexes arise in the proof, see Proposition 3.9 in the next section.

3. Proof of the conjecture on quadratic forms

In this section and the next I will discuss two proofs of Milnor's conjecture on quadratic forms. The first is from [**OVV**], the second was announced in [**M2**]. Both depend on Voevodsky's proof of the norm residue conjecture. As I keep saying, I'm only going to give a vague outline of how the proofs go, but with references for where to find more information on various aspects. The present section deals with the [**OVV**] proof.

3.1. Preliminaries. Recall that we are concerned with the map $\nu\colon K_*^M(F)/2 \to \mathrm{Gr}_I\, W(F)$ defined by $\nu(\{a_1,\dots,a_n\}) = \langle 1,-a_1\rangle \cdots \langle 1,-a_n\rangle$. The fact that I is additively generated by the forms $\langle 1,x\rangle$ shows that ν is obviously surjective; so our task is to prove injectivity. In general, the product $\langle 1,b_1\rangle \cdots \langle 1,b_n\rangle$ is called an ***n*-fold Pfister form**, and denoted $\langle\langle b_1,\dots,b_n\rangle\rangle$. Note that it has dimension 2^n. The proof is intimately tied up with the study of such forms.

Milnor proved that the map $\nu\colon K_2^M(F)/2 \to I^2/I^3$ is an isomorphism. He used ideas of Delzant [**De**] to define Stiefel-Whitney invariants for quadratic forms, which in dimension 2 give a map $I^2/I^3 \to K_2^M(F)/2$. One could explicitly check that this was an inverse to ν. Unfortunately, this last statement generally fails in larger dimensions; the Stiefel-Whitney invariants don't carry enough information. See [**Mr2**, 4.1, 4.2].

3.2. The Orlov-Vishik-Voevodsky proof. We first need to recall some results about Pfister forms proven in the 70's. The first is an easy corollary of the so-called Main Theorem of Arason-Pfister (cf. [**S1**, 4.5.6]). For a proof, see [**EL**, pp. 192-193].

PROPOSITION 3.3 (Elman-Lam). *$\langle\langle a_1,\dots,a_n\rangle\rangle \equiv \langle\langle b_1,\dots,b_n\rangle\rangle$ (mod I^{n+1}) if and only if $\langle\langle a_1,\dots,a_n\rangle\rangle = \langle\langle b_1,\dots,b_n\rangle\rangle$ in $GW(F)$.*

Combining the result for $n = 2$ with Milnor's theorem that $K_2^M(F) \to I^2/I^3$ is an isomorphism, we get the following (note that the minus signs are there because $\nu(\{a_1,\dots,a_n\}) = \langle\langle -a_1,\dots,-a_n\rangle\rangle$):

COROLLARY 3.4. *$\langle\langle a_1,a_2\rangle\rangle = \langle\langle b_1,b_2\rangle\rangle$ in $GW(F)$ if and only if $\{-a_1,-a_2\} = \{-b_1,-b_2\}$ in $K_*^M(F)/2$.*

Say that two n-fold Pfister forms $A = \langle\langle a_1,\dots,a_n\rangle\rangle$ and $B = \langle\langle b_1,\dots,b_n\rangle\rangle$ are **simply-*p*-equivalent** if there are two indices i, j where $\langle\langle a_i,a_j\rangle\rangle = \langle\langle b_i,b_j\rangle\rangle$ and $a_k = b_k$ for all $k \notin \{i,j\}$. The forms A and B are **chain-*p*-equivalent** if there is a chain of forms starting with A and ending with B in which every link of the chain is a simple-p-equivalence. Note that it follows immediately from the previous corollary that if A and B are chain-p-equivalent then $\{-a_1,\dots,-a_n\} = \{-b_1,\dots,-b_n\}$.

The following result is [**EL**, Main Theorem 3.2]:

PROPOSITION 3.5. *Let $A = \langle\langle a_1,\dots,a_n\rangle\rangle$ and $B = \langle\langle b_1,\dots,b_n\rangle\rangle$. The following are equivalent:*

(a) A and B are chain-p-equivalent.
(b) $\{-a_1,\dots,-a_n\} = \{-b_1,\dots,-b_n\}$ in $K_*^M(F)/2$.
(c) $A \equiv B$ (mod I^{n+1}).
(d) $A = B$ in $GW(F)$.

Note that $(a) \Rightarrow (b) \Rightarrow (c)$ is trivial, and $(c) \Rightarrow (d)$ was mentioned above. So the new content is in $(d) \Rightarrow (a)$. I will not give the proof, but refer the reader to [**S1**, 4.1.2]. The result below is a restatement of $(c) \Rightarrow (b)$:

COROLLARY 3.6. *The equality* $\nu(\{a_1, \dots, a_n\}) = \nu(\{b_1, \dots, b_n\})$ *can only occur if* $\{a_1, \dots, a_n\} = \{b_1, \dots, b_n\}$.

Unfortunately the above corollary does not show that ν is injective, as a typical element $x \in K_*^M(F)/2$ is a *sum* of terms $\{a_1, \dots, a_n\}$. A term $\{a_1, \dots, a_n\}$ is called a **pure symbol**, whereas a general $x \in K_*^M(F)$ is just a **symbol**. The key ingredient needed from [**OVV**] is the following:

PROPOSITION 3.7. *If* $x \in K_*^M(F)/2$ *is a nonzero element then there is a field extension* $F \hookrightarrow F'$ *such that the image of* x *in* $K_*^M(F')/2$ *is a nonzero pure symbol.*

It is easy to see that the previous two results prove the injectivity of ν. If $x \in K_n^M(F)/2$ is a nonzero element in the kernel of ν, then by passing to F' we find a nonzero pure symbol which is also in the kernel. Corollary 3.6 shows this to be impossible, however.

We are therefore reduced to proving Proposition 3.7. If we write $x = \underline{a}_1 + \dots + \underline{a}_k$, where each $\underline{a}_i$ is a pure symbol, then we know we can make $\underline{a}_i$ vanish by passing to the function field $F(Q_{\underline{a}_i})$ (where $Q_{\underline{a}_i}$ is the splitting variety produced in the last section). Our goal will be to show that $\underline{a}_i$ is the *only* term that vanishes:

PROPOSITION 3.8 (Orlov-Vishik-Voevodsky). *If* $\underline{a} = \{a_1, \dots, a_n\}$ *is nonzero in* $K_n^M(F)/2$*, then the kernel of* $K_n^M(F)/2 \to K_n^M(F(Q_{\underline{a}}))/2$ *is precisely* $\mathbb{Z}/2$ *(generated by* $\underline{a}$*).*

Granting this for the moment, let i be the largest index for which x is nonzero in $K_n^M(F')/2$, where $F' = F(Q_{\underline{a}_1} \times \dots \times Q_{\underline{a}_i})$. Since x will become zero over $F'(Q_{\underline{a}_{i+1}})$, the above result says that $x = \underline{a}_{i+1}$ in $K_n^M(F')/2$. This is precisely what we wanted.

So finally we have reduced to the same kind of problem we tackled in the last section, namely controlling the map $K_n^M(F)/2 \to K_n^M(F(Q_{\underline{a}}))/2$. The techniques needed to prove Proposition 3.8 are exactly the same as those from the last section. There is a again a homotopical ingredient and a geometric ingredient.

PROPOSITION 3.9. *If* X *is a smooth scheme over* F*, then for every* $n \geq 0$ *there is an exact sequence of the form*

$$0 \to H^{n,n-1}(\check{C}X; \mathbb{Z}/2) \to H^{n,n}(\operatorname{Spec} F; \mathbb{Z}/2) \to H^{n,n}(\operatorname{Spec} F(X); \mathbb{Z}/2).$$

Recall that $H^{n,n}(\operatorname{Spec} E; \mathbb{Z}/2) \cong K_n^M(E)/2$ for any field E. So the above sequence is giving us control over the kernel of $K_*^M(F)/2 \to K_*^M(F(Q_{\underline{a}}))/2$. The proof uses the conclusion from Proposition 2.4 (which is known by Voevodsky's proof of the Milnor conjecture) and some standard manipulations with motivic cohomology. See [**OVV**, Prop. 2.3].

REMARK 3.10. In some sense Proposition 3.9 explains why Čech complexes are destined to come up in the proofs of these conjectures.

If the above proposition is thought of as a 'homotopical' part of the proof, the geometric part is the following. It is deduced using Rost's results on the motive of $Q_{\underline{a}}$; see [**OVV**, Prop. 2.5].

PROPOSITION 3.11. *There is a surjection* $\mathbb{Z}/2 \to H^{2^n-1,2^{n-1}-1}(\check{C}Q_{\underline{a}};\mathbb{Z}/2)$.

The previous two results immediately yield a proof of 3.8. By Proposition 3.9 we must show that $H^{n,n-1}(\check{C}Q_{\underline{a}};\mathbb{Z}/2) \cong \mathbb{Z}/2$ (and we know the group is nontrivial). But we saw in the last section that $H^{n,n-1}(\check{C}Q_{\underline{a}};\mathbb{Z}/2) \cong H^{n+1,n-1}(\tilde{C}Q_{\underline{a}};\mathbb{Z}/2)$, where $\tilde{C}Q_{\underline{a}}$ is the homotopy cofiber of $(\check{C}Q_{\underline{a}})_+ \to (\operatorname{Spec} F)_+$. We also saw that the operation $Q_{n-2}\cdots Q_2Q_1$ gives a monomorphism $H^{n+1,n-1}(\tilde{C}Q_{\underline{a}};\mathbb{Z}/2) \hookrightarrow H^{2^n-1,2^{n-1}-1}(\tilde{C}Q_{\underline{a}};\mathbb{Z}/2)$. But now we are done, since by 3.11 the latter group has at most two elements.

This completes the proof of the injectivity of ν.

4. Quadratic forms and the Adams spectral sequence

In [**M2**] Morel announced a proof of the quadratic form conjecture over characteristic zero fields, using the motivic Adams spectral sequence. The approach depends on having computed the motivic Steenrod algebra, but I'm not sure what the status of this is—certainly no written account is presently available. Despite this frustrating point, Morel's proof is very exciting; while it uses Voevodsky's computation of $H^{*,*}(\operatorname{Spec} F;\mathbb{Z}/2)$—see Remark 2.10—it somehow avoids using any other deep results about quadratic forms! So I'd like to attempt a sketch.

The arguments below take place in the motivic stable homotopy category. All the reader needs to know as background is that it formally behaves much as the usual stable homotopy category, and that there is a bigraded family of spheres $S^{p,q}$. The suspension (in the triangulated category sense) of $S^{p,q}$ is $S^{p+1,q}$, and $S^{2,1}$ is the suspension spectrum of the variety $\mathbb{P}^1$.

4.1. Outline. We have our maps $\nu_n\colon K_n^M(F)/2 \to I^n/I^{n+1}$, and need to prove that they are injective. We will see that the Adams spectral sequence machinery gives us, more or less for free, maps $s_n\colon I^n/I^{n+1} \to K_n^M(F)/(2,J)$ where J is a subgroup of boundaries from the spectral sequence. The composite $s_n\nu_n$ is the natural projection, and so the whole game is to show that J is zero. That is, one needs to prove the vanishing of a line of differentials. Using the multiplicative structure of the spectral sequence and the algebra of the E_2-term, this reduces just to proving that the differentials on a certain 'generic' element vanish. This allows one to reduce to the case of the prime field $\mathbb{Q}$, then to $\mathbb{R}$, and ultimately to a purely topological problem.

4.2. Basic setup. Now I'll expand on this general outline. The first step is to produce a map $q\colon GW(F) \to \{S^{0,0}, S^{0,0}\}$ where $\{-,-\}$ denotes maps in the motivic stable homotopy category. Recall from Section 1.2 that one knows a complete description of $GW(F)$ in terms of generators and relations. For $a \in F^*$ we let $q(\langle a\rangle)$ be the map $\mathbb{P}^1 \to \mathbb{P}^1$ defined in homogeneous coordinates by $[x, y] \to [x, ay]$. By writing down explicit $\mathbb{A}^1$-homotopies one can verify that the relations in $GW(F)$ are satisfied in $\{S^{0,0}, S^{0,0}\}$, and so q extends to a well-defined map of abelian groups. It is actually a ring map. Further details about all this are given in [**M3**].

Now we build an Adams tower for $S^{0,0}$ based on the motivic cohomology spectrum $H\mathbb{Z}/2$. Set $W_0 = S^{0,0}$, and define W_1 by the homotopy fiber sequence $W_1 \to S^{0,0} \to H\mathbb{Z}/2$. Then consider the map $W_1 \cong S^{0,0} \wedge W_1 \to H\mathbb{Z}/2 \wedge W_1$, and let W_2 be the homotopy fiber. Repeat the process to define W_3, W_4, etc. This gives us a tower of cofibrations

$$\begin{tikzcd}
 & H \wedge W_2 & H \wedge W_1 & H \wedge W_0 \\
\cdots \arrow[r] & W_2 \arrow[u] \arrow[r] & W_1 \arrow[u] \arrow[r] & W_0, \arrow[u]
\end{tikzcd}$$

where we have written H for $H\mathbb{Z}/2$. For any Y the tower yields a filtration on $\{Y, S^{0,0}\}$ by letting $\mathcal{F}^n$ be the subgroup of all elements in the image of $\{Y, W_n\}$ (note that there is no *a priori* guarantee that the filtration is Hausdorff.) The tower yields a homotopy spectral sequence whose abutment has something to do with the associated graded of the groups $\{S^{*,0} \wedge Y, S^{0,0}\}$. If the filtration is not Hausdorff these associated graded groups may not be telling us much about $\{S^{*,0} \wedge Y, S^{0,0}\}$, but this will not matter for our application. We will be interested in the case $Y = S^{0,0}$.

Set $E_1^{a,b} = \{S^{a,0}, H \wedge W_b\}$, so that $d_r\colon E_r^{a,b} \to E_r^{a-1,b+r}$. My indexing has been chosen so that the picture of the spectral sequence has $E_1^{a,b}$ in spot (a,b) on a grid, rather than at spot $(b-a,a)$ as is more typical for the Adams spectral sequence—but the picture itself is the same in the end. Formal considerations give inclusions

$$\mathcal{F}^k\{S^{n,0}, S^{0,0}\}/\mathcal{F}^{k+1}\{S^{n,0}, S^{0,0}\} \hookrightarrow E_\infty^{n,k}$$

(however, there is no a priori reason to believe the map is surjective). In particular, if $\mathcal{F}^*$ is the filtration on $\{S^{0,0}, S^{0,0}\}$ then we have inclusions $\mathcal{F}^k/\mathcal{F}^{k+1} \hookrightarrow E_\infty^{0,k}$.

Let $G\mathbb{I}(F)$ be the kernel of the mod 2 dimension function $\dim\colon GW(F) \to \mathbb{Z}/2$. The powers $G\mathbb{I}(F)^n$ define a filtration on $GW(F)$. One can check that q maps $G\mathbb{I}^1$ into $\mathcal{F}^1$. Since the Adams filtration $\mathcal{F}^n$ on $\pi_{0,0}(S^{0,0})$ will be multiplicative, one finds that q maps $G\mathbb{I}^n$ into $\mathcal{F}^n$. So we get maps $(G\mathbb{I})^n/(G\mathbb{I})^{n+1} \to \mathcal{F}^n/\mathcal{F}^{n+1} \to E_\infty^{0,n}$.

In a moment I'll say more about what the Adams spectral sequence looks like in this case, but first let's relate $G\mathbb{I}$ to what we really care about. One easily checks that $G\mathbb{I} = GI \oplus \mathbb{Z}$, where the $\mathbb{Z}$ is the subgroup generated by $\langle 1,1\rangle = 2\langle 1\rangle$. So $G\mathbb{I}^n = GI^n \oplus \mathbb{Z}$, where the $\mathbb{Z}$ is generated by $2^n\langle 1\rangle$. It follows that $G\mathbb{I}^n/G\mathbb{I}^{n+1} \cong [GI^n/GI^{n+1}] \oplus \mathbb{Z}/2$. Finally, recall from Remark 1.5 that the natural map $GI \to I$ is an isomorphism. Putting everything together, we have produced invariants $[I^n/I^{n+1}] \oplus \mathbb{Z}/2 \to E_\infty^{0,n}$.

4.3. Analysis of the spectral sequence. So far the discussion has been mostly formal. We have produced a spectral sequence, but not said anything concrete about it. The usefulness of the above invariants hinges on what $E_\infty^{0,n}$ looks like. If things work as in ordinary topology, then the E_2 term will turn out to be $E_2^{a,b} = \mathrm{Ext}^b_{H^{**}H}(\Sigma^{b+a,0}H^{**}, H^{**})$ where I've again written $H = H\mathbb{Z}/2$ and $\Sigma^{k,0}$ denotes a grading shift on the bi-graded module H^{**}. So we need to know the algebra $H^{**}H$, but unfortunately there is no published source for this calculation. In [**V2**] Voevodsky defines Steenrod operations and shows that they satisfy analogs of the usual Adem relations; he doesn't show that these generate all of $H^{**}H$, though. However, let's assume we knew this—so we are assuming $H^{**}H$ is the algebra Voevodsky denotes A^{**} and calls the motivic Steenrod algebra [**V2**, Section 11].

The form of $H^{**}H$ is very close to that of the usual Steenrod algebra, and so one has a chance at doing some of the Ext computations. In fact, it is not very hard. Some hints about this are given in Appendix B, but for now let me just tell you the important points:

(1) $E_2^{p,q} = 0$ if $p < 0$.
(2) $E_2^{0,0} = \mathbb{Z}/2$.
(3) For $n \geq 1$, $E_2^{0,n} = H^{n,n} \oplus \mathbb{Z}/2$. The inclusion $\oplus_n H^{n,n} \hookrightarrow \oplus_n E_2^{0,n}$ is a ring homomorphism, where the domain is regarded as a subring of H^{**}.

Most of these computations make essential use of Remark 2.10, and therefore depend on Voevodsky's proof of the norm residue conjecture. Also note the connection between (3) and Milnor K-theory, given by the isomorphism $H^{n,n} \cong K_n^M(F)/2$.

The above two facts show that everything in $E_2^{0,n}$ is a permanent cycle and thus $E_\infty^{0,n} = (\mathbb{Z}/2 \oplus K_n^M(F)/2)/J$ where J is the subgroup of all boundaries. Recall that one has maps

$$K_n^M(F)/2 \xrightarrow{\nu_n} I^n/I^{n+1} \to E_\infty^{0,n} \cong [K_n^M(F)/2 \oplus \mathbb{Z}/2]/J.$$

The composition can be checked to be the obvious one. To prove that ν_n is injective, we need to prove that $J = 0$. That is, we need to prove the vanishing of all differentials landing in $E^{0,*}$ (which necessarily come from $E^{1,*}$). As for the computation of the $E^{1,*}$ column, here are the additional facts we need:

(4) $E_2^{1,0} = 0$.
(5) $E_2^{1,1} = H^{0,1} \oplus H^{2,2} \cong \mathbb{Z}/2 \oplus H^{2,2}$.
(6) The images of the two maps

$$E_2^{0,1} \otimes E_2^{1,n-1} \to E_2^{1,n} \qquad E_2^{1,n-1} \otimes E_2^{0,1} \to E_2^{1,n}$$

generate $E_2^{1,n}$ as an abelian group.
(7) The composite $H^{1,1} \otimes H^{2,2} \hookrightarrow E_2^{0,1} \otimes E_2^{1,1} \to E_2^{1,2}$ is zero.

Again, let me say that none of these computations is particularly difficult, and the reader can find some hints in Appendix B. Portions of columns 0 and 1 of our E_2-term are shown below:

$H^{4,4} \oplus \mathbb{Z}/2$	??		
$H^{3,3} \oplus \mathbb{Z}/2$	??		
$H^{2,2} \oplus \mathbb{Z}/2$	??		
$H^{1,1} \oplus \mathbb{Z}/2$	$H^{2,2} \oplus \mathbb{Z}/2$		
$\mathbb{Z}/2$	0		

REMARK 4.4. If one only looks at the $\mathbb{Z}/2$'s appearing in the above diagram, the picture looks just like the ordinary topological Adams spectral sequence. The $\mathbb{Z}/2$'s in our 0th column indeed turn out to be "h_0^n's", just as in topology. The $\mathbb{Z}/2$ in $E_2^{1,1}$ is a little more complicated, though—it doesn't just come from Sq^2, like the usual h_1 does (see Appendix B for what it *does* come from).

We need to prove that all the differentials leaving the $E^{1,*}$ column vanish. By fact (6) and the multiplicative structure of the spectral sequence, it is sufficient to prove that all differentials leaving $E_2^{1,1}$ vanish (starting with $d_2 \colon E_2^{1,1} \to E_2^{0,3}$). We will do this in several steps.

The following result basically shows that, just as in ordinary topology, all the $\mathbb{Z}/2$'s in column 0 survive to E_∞.

LEMMA 4.5. *The image of $d_r \colon E_r^{1,1} \to E_r^{0,r+1}$ lies in the subgroup $H^{r+1,r+1}$, for every $r \geq 2$.*

PROOF. Suppose there is an element $x \in E_r^{1,1}$ such that $d_r(x)$ does not lie in $H^{r+1,r+1}$ (or rather its image in E_r). We can write $x = \underline{a} + y$ where $\underline{a} \in H^{2,2} = K_2^M(F)/2$ and $y \in H^{0,1} \cong \mathbb{Z}/2$. In expressing $\underline{a}$ as a sum of pure symbols, one notes that only a finite number of elements of F are involved. By naturality of the spectral sequence, we can therefore assume F is a finitely-generated extension of $\mathbb{Q}$.

But now we can choose an embedding $F \hookrightarrow \mathbb{C}$, and again use naturality. The groups $K_n^M(\mathbb{C})/2$ are all zero, and therefore our assumption implies that over $\mathbb{C}$ we have $E_{r+1}^{0,r+1} = 0$ (in other words, the $\mathbb{Z}/2$ in $E_2^{0,r+1}$ dies in the spectral sequence). But there is a 'topological realization map' from our spectral sequence over $\mathbb{C}$ to the usual Adams spectral sequence in topology, where we know that none of the $\mathbb{Z}/2$'s in $E^{0,*}$ ever die. $\square$

REMARK 4.6. There is also a purely algebraic proof of the above result. One reduces via naturality to the case of algebraically closed fields, where all the $H^{n,n}$'s are zero. Then one shows that the $\mathbb{Z}/2$'s in the 0th column form a polynomial algebra, and that the composite $\mathbb{Z}/2 \otimes \mathbb{Z}/2 \hookrightarrow E_2^{1,1} \otimes E_2^{0,1} \to E_2^{1,2}$ is zero (just as in ordinary topology). The fact that the spectral sequences is multiplicative takes care of the rest.

LEMMA 4.7. *For $\underline{a} \in H^{2,2}$ one has $d_r(\underline{a}) = 0$, for every r.*

PROOF. It follows from facts (3) and (7), together with the multiplicative structure of the spectral sequence, that everything in the image of $d_r \colon H^{2,2} \to H^{r+1,r+1}$ is killed by $H^{1,1}$. This is the key to the proof.

Let $z = d_r(\underline{a})$. Consider the naturality of the spectral sequence for the map $j \colon F \to F(t)$. It follows from the previous paragraph that $j(z) = d_r(j\underline{a})$ is killed by $F(t)^*$. In particular, $\{t\} \cdot j(z) = 0$ in $K_{r+2}^M(F(t))/2$. But by [**Mr2**, Lem. 2.1] there is a map $\partial_t \colon K_{r+2}^M(F(t))/2 \to K_{r+1}^M(F)/2$ with the property that $\partial_t(\{t\} \cdot j(z)) = z$. So we conclude that $z = 0$, as desired. $\square$

PROPOSITION 4.8. *All differentials leaving $E^{1,1}$ are zero.*

PROOF. Recall $E_2^{1,1} \cong H^{0,1} \oplus H^{2,2} \cong \mathbb{Z}/2 \oplus H^{2,2}$. By the previous lemma we are reduced to analyzing the maps $d_r \colon H^{0,1} \to H^{r+1,r+1}$. Since $H^{0,1}(\mathbb{Q}) \to H^{0,1}(F)$ is an isomorphism, it suffices to prove the result in the case $F = \mathbb{Q}$.

Now use naturality with respect to the field extension $\mathbb{Q} \hookrightarrow \mathbb{R}$. The maps $K_n^M(\mathbb{Q})/2 \to K_n^M(\mathbb{R})/2$ are isomorphisms for $n \geq 3$ (see Appendix A), so now we've reduced to $F = \mathbb{R}$. But here we can again use a 'topological realization' map to compare our Adams spectral sequence to the corresponding one in the context of $\mathbb{Z}/2$-equivariant homotopy theory. This map is readily seen to be an isomorphism on the $E^{0,*}$ column: the point is that the $\mathbb{Z}/2$-equivariant cohomology groups $H^{n,n}$ are isomorphic to the corresponding mod 2 motivic cohomology groups over $\mathbb{R}$ (see [**Du**, 2.8, 2.11], for instance). We are essentially seeing a reflection of the fact that $GW(\mathbb{R})$ may be identified with the Burnside ring of $\mathbb{Z}/2$, which coincides with $\{S^{0,0}, S^{0,0}\}$ in the $\mathbb{Z}/2$-equivariant stable homotopy category. In any case, we are finally reduced to showing the vanishing of certain differentials in a topological Adams spectral sequence: the paper [**LZ**] seems to essentially do this (but I haven't thought about this part carefully—I'm relying on remarks from [**M2**]). $\square$

This completes Morel's proof of the quadratic form conjecture for characteristic zero fields (modulo the identification of $H^{**}H$, which we assumed).

REMARK 4.9. We restricted to characteristic zero fields because the identification of $H^{**}H$ has never been claimed in characteristic p. If we make the wild guess that in positive characteristic $H^{**}H$ still has the same form, most of the argument goes through verbatim. There are two exceptions, where we used topological realization functors. The first place was to show that the image of the d_r's didn't touch the $\mathbb{Z}/2$'s in $E_2^{0,*}$, but Remark 4.6 mentioned that this could be done another way. The second place we used topological realization was at the final stage of the argument, to analyze the differentials $d_r\colon H^{0,1} \to H^{r+1,r+1}$. As before, this reduces to the case of a prime field. But for F a finite field one has $K_n^M(F) = 0$ for $n \geq 2$, so for prime fields there is in fact nothing to check.

In summary, the same general argument would work in characteristic p if one knew that $H^{**}H$ had the same form.

4.10. Further reading. There is very little completed literature on the subjects discussed in this section. Several documents are available on Morel's website, however; the draft [**M5**] is particularly relevant, although it only slightly expands on [**M2**]. For information on the motivic Steenrod algebra, see [**V2**]. Finally, Morel recently released another proof of Milnor's quadratic form conjecture, using very different methods. See [**M4**].

1. Some examples of the Milnor conjectures

This is a supplement to Section 1. We examine the Milnor conjectures in the cases of certain special fields F.

(a) F is algebraically closed. Since $F = F^2$, every nondegenerate form is isomorphic to one of the form $\langle 1, 1, \ldots, 1\rangle$. So $GW(F) \cong \mathbb{Z}$, and $W(F) \cong \mathbb{Z}/2$ with $I(F) = 0$. Thus, $\mathrm{Gr}_I\, W(F) \cong \mathbb{Z}/2$.

The absolute Galois group is trivial, so $H^*(F; \mathbb{Z}/2) = \mathbb{Z}/2$.

Finally, the fact that $F = F^2$ implies that $K^M_*(F)/2 = 0$ for $* \geq 1$. This is because the generators all lie in $K^M_1(F)$, and if $a = x^2$ then $\{a\} = \{x^2\} = 2\{x\} = 0 \in K^M_1(F)/2$.

(b) $F = F^2$. This case is suggested by the previous one. We only need to check that the hypothesis implies $H^*(F; \mathbb{Z}/2) = 0$ for $* \geq 1$. Strangely, I haven't been able to find an easy proof of this.

(c) $F = \mathbb{R}$. In this case we know forms are classified by their rank and signature, and it follows that $GW(\mathbb{R})$ is the free abelian group generated by $\langle 1\rangle$ and $\langle -1\rangle$. Also, $\langle -1\rangle^2 = \langle 1\rangle$. So $GW(\mathbb{R}) \cong \mathbb{Z}[x]/(x^2 - 1)$, and $W(\mathbb{R}) \cong \mathbb{Z}$ with $I(\mathbb{R}) = 2\mathbb{Z}$. Hence $\mathrm{Gr}_I\, W(\mathbb{R}) \cong \mathbb{Z}/2[a]$.

The absolute Galois group of $\mathbb{R}$ is $\mathbb{Z}/2$, so $H^*(\mathbb{R}; \mathbb{Z}/2) = H^*(\mathbb{Z}/2; \mathbb{Z}/2) = \mathbb{Z}/2[a]$.

Finally we consider $K^M_*(\mathbb{R})/2$. The group $K^M_1(\mathbb{R})/2 = \mathbb{R}^*/(\mathbb{R}^*)^2 \cong \{1, -1\}$ (the set consisting of 1 and -1). A similar calculation, based on the fact that every element of $\mathbb{R}$ is a square up to sign, shows that $K^M_i(\mathbb{R})/2 \cong \mathbb{Z}/2$ for every i, with the nonzero element being $\{-1, -1, \ldots, -1\}$. So $K^M_*(\mathbb{R})/2 \cong \mathbb{Z}/2[a]$ as well.

(d) $F = \mathbb{F}_q$, q odd. Here $F^* \cong \mathbb{Z}/(q-1)$ and so $K^M_1/2 = F^*/(F^*)^2 \cong \mathbb{Z}/2$. If g is the generator, then $\{g, g, \ldots, g\}$ generates $K^M_n/2$ (but may be zero). In fact one can show (cf. [**Mr2**, Ex. 1.5]) that $\{g, g\} = 0$ in K^M_2, from which it follows that $K^M_* = 0$ for $* \geq 2$. So $K^M_*(F)/2 \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$, in degrees 0 and 1.

For a finite field the absolute Galois group is $\hat{\mathbb{Z}}$, the profinite completion of $\mathbb{Z}$. The Galois cohomology $H^*(\hat{\mathbb{Z}}; \mathbb{Z}/2)$ is just the mod 2 cohomology of $B\mathbb{Z} \simeq S^1$; so it is $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, with the generators in degrees 0 and 1.

Again noting that $F^*/(F^*)^2 \cong \mathbb{Z}/2$, it follows that the Grothendieck-Witt group is generated by $\langle 1\rangle$ and $\langle g\rangle$. A simple counting argument (cf. [**S1**, Lem. 2.3.7]) shows that every element of $\mathbb{F}_q^*$ is a sum of two squares. Writing $g = a^2 + b^2$ one finds that

$$\langle 1, 1\rangle = \langle a^2, b^2\rangle = \langle a^2 + b^2, a^2b^2(a^2 + b^2)\rangle = \langle a^2 + b^2, a^2 + b^2\rangle = \langle g, g\rangle.$$

That is, $2(\langle 1\rangle - \langle g\rangle) = 0$. It follows that $GW(F) = \mathbb{Z} \oplus \mathbb{Z}/2$, with corresponding generators $\langle 1\rangle$ and $\langle 1\rangle - \langle g\rangle$.

The computation of the Witt group depends on whether or not -1 is a square; since $F^* = \mathbb{Z}/(q-1)$ and -1 has order 2, then -1 is a square precisely when $4|(q-1)$. So if $q \equiv 1 (\mathrm{mod}\ 4)$ then $\langle 1\rangle = \langle -1\rangle$ and $W(F) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$; in this case $I(F) = (\langle 1\rangle - \langle g\rangle) \cong \mathbb{Z}/2$. If $q \equiv 3 (\mathrm{mod}\ 4)$ then $\langle g\rangle = \langle -1\rangle$ and we have $W(F) \cong \mathbb{Z}/4$ with $I(F) = (2)$. In either case $\mathrm{Gr}_I\, W(F) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$.

REMARK 1.1. Although Milnor's quadratic form conjecture says that $\operatorname{Gr}_I W(F)$ depends only on the absolute Galois group of F, this example makes it clear that the same cannot be said for $W(F)$ itself.

(e) $F = \mathbb{Q}$. This case is considerably harder, so we will only make a few observations. Note that as an abelian group one has

$$\mathbb{Q}^* \cong \mathbb{Z}/2 \times \Big(\oplus_p \mathbb{Z}\Big),$$

by the fundamental theorem of arithmetic; the direct sum is over the set of all primes. Here the isomorphism sends a fraction q to its sign (in the $\mathbb{Z}/2$ factor) together with the list of exponents in the prime factorization of q. So $K_1^M(\mathbb{Q})/2 \cong \mathbb{Z}/2 \oplus (\oplus_p \mathbb{Z}/2)$.

As the above isomorphism may suggest, to go further it becomes convenient to work with one completion at a time. The case $F = \mathbb{R}$ has already been discussed, so what is left is the p-adics. We will return to $F = \mathbb{Q}$ after discussing them.

(f) $F = \mathbb{Q}_p$. We will concentrate on the case where p is odd; the case $p = 2$ is similar, and can be left to the reader. We know $K_1^M(\mathbb{Q}_p)/2 \cong H^1(\mathbb{Q}_p; \mathbb{Z}/2) \cong \mathbb{Q}_p^*/(\mathbb{Q}_p^*)^2$. A little thought (cf. [**S1**, 5.6.2]) shows this group is $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, with elements represented by 1, g, p, and pg, where $1 < g < p$ is any integer which generates the multiplicative group $\mathbb{F}_p^*$. By [**Se3**, Section II.5.2] one has $H^2(\mathbb{Q}_p; \mathbb{Z}/2) \cong \mathbb{Z}/2$ and $H^i(\mathbb{Q}_p; \mathbb{Z}/2) = 0$ for $i \geq 3$.

The fact that $K_1^M(\mathbb{Q}_p)/2$ only has four elements tells us that $K_*^M(\mathbb{Q}_p)/2$ can't be too big. By finding the appropriate relations to write down, Calvin Moore proved that $K_*^M(\mathbb{Q}_p)/2 = 0$ for $* \geq 3$ [**Mr2**, Ex. 1.7], and that $K_2^M(\mathbb{Q}_p)/2 = \mathbb{Z}/2$. This is an exercise for the reader.

The group $GW(\mathbb{Q}_p)$ will be generated by the four elements $\langle 1 \rangle$, $\langle g \rangle$, $\langle p \rangle$, and $\langle pg \rangle$. The theory again depends on whether or not -1 is a square, which is when $p \equiv 1 (\text{mod } 4)$. When $p \equiv 1 (\text{mod } 4)$ one has $\langle 1 \rangle = \langle -1 \rangle$ and so $\langle x \rangle = \langle -x \rangle$ for any x. As a result $\langle g, g \rangle = \langle g, -g \rangle = \langle 1, -1 \rangle = \langle 1, 1 \rangle$, and similarly $\langle p, p \rangle = \langle pg, pg \rangle = \langle 1, 1 \rangle$. One finds that $GW(\mathbb{Q}_p) = \mathbb{Z} \oplus (\mathbb{Z}/2)^3$ with corresponding generators $\langle 1 \rangle$, $\langle 1 \rangle - \langle p \rangle$, $\langle 1 \rangle - \langle g \rangle$, and $\langle 1 \rangle - \langle pg \rangle$. Since $\langle 1, -1 \rangle = 2\langle 1 \rangle$, $W(\mathbb{Q}_p) = (\mathbb{Z}/2)^4$ with the same generators. I is generated by $\langle 1, p \rangle$, $\langle 1, g \rangle$, and $\langle 1, pg \rangle$; I^2 is generated by $\langle 1, p, g, pg \rangle$, and $I^3 = 0$. So $\operatorname{Gr}_I W = \mathbb{Z}/2 \oplus (\mathbb{Z}/2 \oplus \mathbb{Z}/2) \oplus \mathbb{Z}/2$. Note that this is the first example we've seen where $I^2 \neq 2I$.

When $p \equiv 3 (\text{mod } 4)$ we can take $g = -1$. One has $\langle 1, 1 \rangle = \langle -1, -1 \rangle$ by the same reasoning as for $\mathbb{F}_p$ (-1 is the sum of two squares), and so $\langle p, p \rangle = \langle -p, -p \rangle$. Note that

$$\langle p, p, p, p \rangle = \langle p, -p, -p, p \rangle = \langle 1, -1, -1, 1 \rangle = \langle 1, 1, 1, 1 \rangle$$

and so $4(\langle 1 \rangle - \langle p \rangle) = 0$. Also,

$$\langle p, p, p \rangle = \langle p, -p, -p \rangle = \langle 1, -1, -p \rangle \quad \text{and} \quad \langle 1, 1, 1 \rangle = \langle 1, -1, -1 \rangle.$$

So $3(\langle 1 \rangle - \langle p \rangle) = \langle -1 \rangle - \langle -p \rangle$. Of course $GW(\mathbb{Q}_p)$ is generated by $\langle 1 \rangle$, $\langle 1 \rangle - \langle -1 \rangle$, $\langle 1 \rangle - \langle p \rangle$, and $\langle 1 \rangle - \langle -p \rangle$, and the previous computation shows the last generator is not needed. So we have a surjective map $\mathbb{Z} \oplus \mathbb{Z}/2 \oplus \mathbb{Z}/4 \to GW(\mathbb{Q}_p)$ sending the standard generators to $\langle 1 \rangle$, $\langle 1 \rangle - \langle -1 \rangle$, and $\langle 1 \rangle - \langle p \rangle$. This is readily checked to be injective once one knows that $\langle 1, 1 \rangle \ncong \langle p, p \rangle$. If these forms were isomorphic it would follow by reduction mod some power of p that $\langle 1, 1 \rangle$ was isotropic over some $\mathbb{F}_{p^e}$; that is, we would have $\langle 1,1\rangle \cong \langle 1,-1\rangle$. But we've already computed $GW(\mathbb{F}_{p^e})$, and know this is not the case.

The Witt ring is $W(\mathbb{Q}_p) \cong \mathbb{Z}/4 \oplus \mathbb{Z}/4$ with generators $\langle 1\rangle$ and $\langle 1\rangle - \langle p\rangle$. The ideal I is generated by $2\langle 1\rangle$ and $\langle 1\rangle - \langle p\rangle$; I^2 is generated by $2(\langle 1\rangle - \langle p\rangle)$; $I^3 = 0$. Again we have $\operatorname{Gr}_I W \cong \mathbb{Z}/2 \oplus (\mathbb{Z}/2 \oplus \mathbb{Z}/2) \oplus \mathbb{Z}/2$.

(g) Return to $F = \mathbb{Q}$. Our understanding of the higher Milnor K-groups of $\mathbb{Q}$ is based on passing to the various completions $\mathbb{Q}_p$ and $\mathbb{R}$. A computation of Bass and Tate [**Mr2**, Lem. A.1] gives an exact sequence

$$0 \to K_2^M(\mathbb{Q})/2 \to K_2^M(\mathbb{R})/2 \oplus \Big(\oplus_p K_2^M(\mathbb{Q}_p)/2\Big) \to \mathbb{Z}/2 \to 0,$$

and we already know $K_2^M(\mathbb{Q}_p)/2 \cong K_2^M(\mathbb{R})/2 \cong \mathbb{Z}/2$. A computation of Tate [**Mr2**, Th. A.2, Ex. 1.8] shows that for $* \geq 3$ one has

$$K_*^M(\mathbb{Q})/2 \cong \oplus_p K_*^M(\mathbb{Q}_p)/2 \oplus K_*^M(\mathbb{R})/2 \cong 0 \oplus \mathbb{Z}/2.$$

To compute $H^*(\mathbb{Q};\mathbb{Z}/2)$ we again work one completion at a time. A theorem of Tate [**Se3**, Section II.6.3, Th. B] says that for $i \geq 3$ one has

$$H^i(\mathbb{Q};\mathbb{Z}/2) \cong H^i(\mathbb{R};\mathbb{Z}/2) \times \prod_p H^i(\mathbb{Q}_p;\mathbb{Z}/2) \cong H^i(\mathbb{R};\mathbb{Z}/2) \cong \mathbb{Z}/2.$$

Our computation of $\mathbb{Q}^*/(\mathbb{Q}^*)^2 \cong H^1(\mathbb{Q};\mathbb{Z}/2)$ shows that the map $H^1(\mathbb{Q};\mathbb{Z}/2) \to H^1(\mathbb{R};\mathbb{Z}/2) \times \prod_p H^1(\mathbb{Q}_p;\mathbb{Z}/2)$ is injective. More of Tate's work [**Se3**, Sec. II.6.3, Th. A] identifies the dual of the kernel with the kernel of $H^2(\mathbb{Q};\mathbb{Z}/2) \to H^2(\mathbb{R};\mathbb{Z}/2) \times (\oplus_p H^2(\mathbb{Q}_p;\mathbb{Z}/2))$—thus, this latter map is also injective. Using this, [**Se3**, Sec. II.6.3, Th. C] gives a short exact sequence

$$0 \to H^2(\mathbb{Q};\mathbb{Z}/2) \to H^2(\mathbb{R};\mathbb{Z}/2) \oplus (\oplus_p H^2(\mathbb{Q}_p;\mathbb{Z}/2)) \to \mathbb{Z}/2 \to 0.$$

As we have already remarked that $H^2(\mathbb{Q}_p;\mathbb{Z}/2) = H^2(\mathbb{R};\mathbb{Z}/2) = \mathbb{Z}/2$, this completes the calculation of $H^*(\mathbb{Q};\mathbb{Z}/2)$.

The method for computing the Witt group $W(\mathbb{Q})$ proceeds similarly by working one prime at a time. See [**S1**, Section 5.3]. One has an isomorphism of groups $W(\mathbb{Q}) \cong \mathbb{Z} \oplus (\oplus_p W(\mathbb{F}_p))$ [**S1**, Thm. 5.3.4]. With enough trouble one can compute $\operatorname{Gr}_I W(\mathbb{Q})$, but we will leave this for the reader to consider.

Remark 1.2. Note that the verification of the Milnor conjectures for $F = \mathbb{Q}$ tells us exactly how to classify quadratic forms over $\mathbb{Q}$ by invariants. First one needs the invariants over $\mathbb{R}$ (which are just rank and signature), and then one needs the invariants over each $\mathbb{Q}_p$—but for $\mathbb{Q}_p$ one has $I^3 = 0$, and so p-adic forms are classified by the three classical invariants e_0, e_1, and e_2. These observations are essentially the content of the classical Hasse-Minkowski theorem.

The method we've used above, of working one completion at a time, works for all global fields; this is due to Tate for Galois cohomology, and Bass and Tate for K_*^M. In this way one verifies the Milnor conjecture for this class of fields [**Mr2**, Lemma 6.2]. Note in particular that the class includes all finite extensions of $\mathbb{Q}$.

2. More on the motivic Adams spectral sequence

This final section is a supplement to Section 4. I will give some hints on computing the E_2-term of the motivic Adams spectral sequence, for the reader who would like to try this at home. The computations are not hard, but there are several small issues that are worth mentioning.

2.1. Setting things up. $H^{**}H$ is the algebra of operations on mod 2 motivic cohomology. We will write this as $\mathcal{A}$ from now on. There is the Bockstein $\beta \in \mathcal{A}^{1,0}$ and there are squaring operations $Sq^{2i} \in \mathcal{A}^{2i,i}$. We set $Sq^{2i+1} = \beta Sq^{2i} \in \mathcal{A}^{2i+1,i}$. Finally, there is an inclusion of rings $H^{**} \to \mathcal{A}$ sending an element t to the operation left-multiplication-by-t. Under our standing assumptions about $\mathcal{A}$ (see Section 4), it is free as a left H^{**}-module with a basis consisting of the admissible sequences $Sq^{i_1}Sq^{i_2}\cdots Sq^{i_k}$.

There are two main differences between what happens next and what happens in ordinary topology. These are:

(a) The vector space $H^{**} = H^{**}(pt)$, regarded as a left $\mathcal{A}$-module, is nontrivial.
(b) The image of $H^{**} \hookrightarrow \mathcal{A}$ is not central.

The above two facts are connected. Let $t \in H^{**}$ and let Sq denote some Steenrod operation. It is not true in general that $Sq(t \cdot x) = t \cdot Sq(x)$—instead there is a Cartan formula for the left-hand side [**V2**, 9.7], which involves Steenrod operations on t. So the operations $Sq \cdot t$ and $t \cdot Sq$ are not the same element of $\mathcal{A}$. There is one notable exception, which is when all the Steenrod squares vanish on t. This happens for elements in $H^{n,n}$, for dimension reasons. So we have

(c) Every element of $H^{n,n}$ is central in $\mathcal{A}$.

It is important that we can completely understand H^{**} as an $\mathcal{A}$-module. This will follow from (1) the fact that $H^{**} \cong \big(\oplus_n H^{n,n}\big)[\tau]$ (see Remark 2.10); (2) all Steenrod operations vanish on $H^{n,n}$ for dimension reasons; (3) all Sq^i's vanish on τ except for Sq^1, and $Sq^1(\tau) = \rho = \{-1\} \in H^{1,1}$; (4) the Cartan formula. In particular we note the following two facts about H^{**}, which are all that will be needed later (the second fact only needs Remark 2.10):

(d) The map $Sq^2 \colon H^{n-1,n} \to H^{n+1,n+1}$ is zero for all $n \geq 1$.
(e) The map $H^{p,q} \otimes H^{i,j} \to H^{p+i,q+j}$ is surjective for $q \geq p \geq 0$ and $j \geq i \geq 0$.

We are aiming to compute $\mathrm{Ext}^a_{\mathcal{A}}(H^{**}, \Sigma^{b,0}H^{**})$. In ordinary topology we could use the normalized bar construction to do this, but one has to be careful here because H^{**}, as a left $\mathcal{A}$-module, is not the quotient of $\mathcal{A}$ by a two-sided ideal. One way to see this is to use the fact that $Sq^1(\tau) = \rho$. Under the quotient map $\mathcal{A} \to H^{**}$ sending θ to $\theta(1)$, Sq^1 maps to zero but $Sq^1\tau$ does not (it maps to ρ).

So instead of the normalized bar construction we must use the unnormalized one. This can be extremely annoying, but for the most part it turns out not to influence the "low-dimensional" calculations we're aiming for. It is almost certainly an issue when computing past column two of the Adams E_2 term, though. Anyway, let

$$B_n = \mathcal{A} \otimes_{H^{**}} \mathcal{A} \otimes_{H^{**}} \cdots \otimes_{H^{**}} \mathcal{A} \otimes_{H^{**}} H^{**}$$

($n+1$ copies of $\mathcal{A}$). The final H^{**} can be dropped off, of course, but it's useful to keep it there because the $\mathcal{A}$-module structure on H^{**} is nontrivial and enters into the definition of the boundary map. If we denote the generators of B_n as $x = a[\theta_1|\theta_2|\cdots|\theta_n]t$ then the differential is

$$d(x) = (a\theta_1)[\theta_2|\cdots|\theta_n]t + a[\theta_1\theta_2|\theta_3|\cdots|\theta_n]t + \cdots + a[\theta_1|\cdots|\theta_{n-1}]\theta_n(t).$$

The good news is that our coefficients have characteristic 2, and so we don't have to worry about signs. Note that B_n, as a left H^{**}-module, is free on generators $1[\theta_1|\cdots|\theta_n]1$ where each θ_i is an admissible sequence of Steenrod operations (and we must include the possibility of the null sequence $Sq^0 = 1$). We will often drop the 1's off of either end of the bar element, for convenience.

Generators of $\mathrm{Hom}_{\mathcal{A}}(B_n, H^{**})$ can be specified by giving a bar element $[\theta_1|\cdots|\theta_n]$ together with an element $t \in H^{**}$. This data defines a homomorphism $B_n \to H^{**}$ sending the generator $[\theta_1|\cdots|\theta_n]$ to t and all other generators of B_n to zero. Let's denote this homomorphism by $t[\theta_1|\cdots|\theta_n]^*$. These elements generate $\mathrm{Hom}_{\mathcal{A}}(B_n, H^{**})$ as an abelian group.

The last general point to make concerns the multiplicative structure in the cobar construction. If we were working with $\mathrm{Ext}_A(k,k)$ where k is commutative and A is an augmented k-algebra, multiplying two of the above generators in the cobar complex just amounts to concatenating the bar elements—the labels $t \in k$ commute with the θ's, and so can be grouped together: e.g. $t[\theta_1|\cdots|\theta_n]\cdot u[\alpha_1|\cdots|\alpha_k] = tu[\theta_1|\cdots|\theta_n|\alpha_1|\cdots|\alpha_k]$. In our case, the fact that H^{**} is not central in $\mathcal{A}$ immensely complicates the product on the cobar complex: very roughly, the u has to be commuted across each θ_i, and in each case a resulting Cartan formula will introduce new terms into the product. Luckily there is one case where these complications aren't there, which is when $u \in H^{n,n}$—for then u is in the center of $\mathcal{A}$, and the product works just as above. We record this observation for future use:

(f) $t[\theta_1|\cdots|\theta_n]^* \cdot u[\alpha_1|\cdots|\alpha_k]^* = tu[\theta_1|\cdots|\theta_n|\alpha_1|\cdots|\alpha_k]^*$ when $u \in H^{q,q}$.

2.2. Computations. We are trying to compute the groups $\mathrm{Ext}^a_{\mathcal{A}}(H^{**}, \Sigma^{b,0}H^{**})$, and from here on everything is fairly straightforward. As an example let's look at $b=1$. Since $H^{p,q} \neq 0$ only when $0 \le p \le q$, one sees that $\mathrm{Hom}_{\mathcal{A}}(B_0, H^{**}) = 0$ and $\mathrm{Hom}_{\mathcal{A}}(B_1, \Sigma^{1,0}H^{**}) \cong H^{0,0} \oplus H^{1,1}$. The generators for this group are elements of the form $s[Sq^1]^*$ and $t[Sq^2]^*$, where $s \in H^{0,0}$ and $t \in H^{1,1}$.

We likewise find that $\mathrm{Hom}_{\mathcal{A}}(B_2, \Sigma^{1,0}H^{**}) \cong H^{0,1} \oplus H^{0,1} \oplus H^{0,1} \oplus H^{0,1}$, generated by elements $s[Sq^1|1]^*$, $s[1|Sq^1]^*$, $t[Sq^2|1]^*$, and $t[1|Sq^2]^*$. A similar analysis shows that $\mathrm{Hom}_{\mathcal{A}}(B_n, \Sigma^{1,0}H^{**})$ only has such 'degenerate' terms for $n \ge 2$. No degenerate terms like these contribute elements to Ext (at worst they can contribute relations to Ext). So the Ext^n's vanish for $n \ge 2$. An analysis of the coboundary shows that everything in dimension 1 is a cycle. So we find that

$$0 = \mathrm{Ext}^0(H^{**}, \Sigma^{1,0}H^{**}) = \mathrm{Ext}^n(H^{**}, \Sigma^{1,0}H^{**}), \quad \text{for } n \ge 2$$

and

$$\mathrm{Ext}^1(H^{**}, \Sigma^{1,0}H^{**}) \cong H^{0,0} \oplus H^{1,1}$$

with a typical element in the latter group having the form $s[Sq^1]^* + t[Sq^2]^*$ (where $s \in H^{0,0}$ and $t \in H^{1,1}$).

In general, one sees for degree reasons that the 'non-degenerate' terms in $\mathrm{Hom}_{\mathcal{A}}(B_n, \Sigma^{n,0}H^{**})$ all have the form $t[\theta_1|\cdots|\theta_n]^*$ where each θ_i is either Sq^1 or Sq^2. In $\mathrm{Hom}_{\mathcal{A}}(B_{n-1}, \Sigma^{n,0}H^{**})$ one has non-degenerate terms $u[\theta_1|\cdots|\theta_{n-1}]^*$ of the following types:

(i) Each $\theta_i \in \{Sq^1, Sq^2\}$, and at least one Sq^2 occurs. Here $u \in H^{j-1,j}$ where j is the number of Sq^2's.
(ii) Each $\theta_i \in \{Sq^1, Sq^2, Sq^3\}$, and exactly one Sq^3 occurs. Here $u \in H^{j+1,j+1}$ where j is the number of Sq^2's.
(iii) Each $\theta_i \in \{Sq^1, Sq^2, Sq^2Sq^1\}$, and exactly one Sq^2Sq^1 occurs. Here one has $u \in H^{j+1,j+1}$ where j is the number of Sq^2's.
(iv) Each $\theta_i \in \{Sq^1, Sq^2, Sq^4\}$, and exactly one Sq^4 occurs. Here $u \in H^{j+2,j+2}$ where j is the number of Sq^2's.

To analyze the part of the boundary $B_n \to B_{n-1}$ that we care about, one only needs to know the Adem relations $Sq^1Sq^2 = Sq^3$ and $Sq^2Sq^2 = \tau Sq^3Sq^1$. (In fact, since Sq^3Sq^1 doesn't appear in any of the bar elements relevant to $\mathrm{Hom}(B_{n-1}, \Sigma^{n,0}H^{**})$, one may as well pretend $Sq^2Sq^2 = 0$.) From this it's easy to compute that $\mathrm{Ext}^n(H^{**}, \Sigma^{n,0}H^{**}) \cong H^{0,0} \oplus H^{n,n}$ where a typical element has the form $s[Sq^1|Sq^1|\cdots|Sq^1]^* + t[Sq^2|Sq^2|\cdots|Sq^2]^*$. The computation uses remark 2.1(d). Also, one sees that all elements $s[Sq^1|Sq^2]^*$ and $s[Sq^2|Sq^1]^*$ are zero in Ext^2 (being the coboundaries of $s[Sq^3]^*$ and $s[Sq^2Sq^1]^*$, respectively). Using remark (f) from Section 2.1, this completely determines $\oplus_n \mathrm{Ext}^n(H^{**}, \Sigma^n H^{**})$ as a subring of the whole Ext-algebra.

The next step is to compute $\mathrm{Ext}^0(H^{**}, \Sigma^{1,0}H^{*,*})$, $\mathrm{Ext}^1(H^{**}, \Sigma^{2,0}H^{*,*})$, and $\mathrm{Ext}^2(H^{**}, \Sigma^{3,0}H^{*,*})$ completely. The first group is readily seen to vanish. For the second group one has to grind out another term of the bar construction, but it's a very small term. One finds that

$$\mathrm{Ext}^1(H^{**}, \Sigma^{2,0}H^{*,*}) \cong H^{0,1} \oplus H^{2,2}$$

where the generators have the form $s[Sq^2]^* + (Sq^1s)[Sq^3]^*$ and $t[Sq^4]^*$. To get the Ext^2 group one will need three more Adem relations, namely

$$Sq^2Sq^3 = Sq^5 + Sq^4Sq^1, \quad Sq^2Sq^4 = Sq^6 + \tau Sq^5Sq^1, \quad \text{and} \quad Sq^3Sq^2 = \rho Sq^3Sq^1.$$

Then the same kind of coboundary calculations (but a few more of them) show that

$$\mathrm{Ext}^2(H^{**}, \Sigma^{3,0}H^{*,*}) \cong H^{1,2} \oplus H^{2,2}$$

where the generators are $s[Sq^2|Sq^2]^* + (Sq^1s)[Sq^3|Sq^2]^*$ and $t[Sq^1|Sq^4]^* = t[Sq^4|Sq^1]^*$ (these last two classes are the same in Ext). It is important to note that all elements $u[Sq^2|Sq^4]^*$ and $u[Sq^4|Sq^2]^*$ are coboundaries (of $u[Sq^6]^*$ and $u[Sq^4Sq^2]^*$, respectively). This justifies fact (7) on page 20. To justify fact (6) from that same page (for $n = 2$), one notices that the cycles $s[Sq^2|Sq^2]^* + (Sq^1s)[Sq^3|Sq^2]^*$ and $t[Sq^4|Sq^1]^*$ decompose as a products

$$\big(s_1[Sq^2]^* + (Sq^1s_1)[Sq^3]^*\big) \cdot (s_2[Sq^2]^*) \qquad \text{and} \qquad \big(t_1[Sq^4]^*\big) \cdot \big(t_2[Sq^1]^*\big)$$

for some $s_1 \in H^{0,1}$, $s_2 \in H^{1,1}$, $t_1 \in H^{2,2}$, and $t_2 \in H^{0,0}$. This uses remarks (e) and (f) from Section 2.1, together with the fact that $(Sq^1s_1)s_2 = Sq^1(s_1s_2)$ for $s_2 \in H^{2,2}$ (by the Cartan formula).

The final step is to analyze the groups $\mathrm{Ext}^{n-1}(H^{**}, \Sigma^{n,0}H^{**})$ for $n \geq 4$; these complete the $E^{1,*}$ column of the Adams spectral sequence. One doesn't have to compute them explicitly, just enough to know that every element is decomposable as a sum of products from $\mathrm{Ext}^{n-2}(H^{**}, \Sigma^{n-1,0}H^{**})$ and $\mathrm{Ext}^1(H^{**}, \Sigma^{1,0}H^{**})$.

The calculations involve nothing more than what we've done so far, except for more sweat. It's fairly easy to write down all the cocycles made up from the classes of types (i)-(iv) listed previously. All bar elements which have a Sq^4 in them are cocycles, for instance. But note that such a bar element will either begin or end with a Sq^1 or a Sq^2, so that it decomposes as a product of smaller degree cocycles (this again depends on 2.1(e,f)). One also finds cocycles of the form

$$s[Sq^1|Sq^1|\cdots|Sq^3|Sq^1|\cdots|Sq^1]^* + s[Sq^1|Sq^1|\cdots|Sq^2Sq^1|Sq^1|\cdots|Sq^1]^*,$$

but for each of these a common $[Sq^1]^*$ can be pulled off of either the left or right side—again showing it to be decomposable.

Certainly there are cocycles which are *not* decomposable, like ones of the form

$$s[Sq^2|Sq^1|\cdots|Sq^1|Sq^3]^* + s[Sq^2Sq^1|Sq^1|\cdots|Sq^1|Sq^2]^*.$$

But this is the coboundary of $s[Sq^2Sq^1|Sq^1|\cdots|Sq^1|Sq^3]$, and so vanishes in Ext.

Anyway, I am definitely not going to give all the details. But with enough diligence one can see that all elements of $\mathrm{Ext}^{n-1}(H^{**}, \Sigma^{n,0}H^{**})$ for $n \geq 3$ do indeed decompose into products.

REMARK 2.3. A final note about Adem relations, for those who want to try their hand at further calculations. Every formula I've seen for the motivic Adem relations—in publications or preprints—seems to either contain typos or else is just plain wrong. A good test for a given formula is to see whether it gives $Sq^3Sq^2 = \rho Sq^3Sq^1$ (this formula follows from the smaller Adem relation $Sq^2Sq^2 = \tau Sq^3Sq^1$, the derivation property of the Bockstein, the fact that $\beta^2 = 0$, and the identity $Sq^3 = \beta Sq^2$).

Bibliography

[A] L. Ahlfors, *Complex analysis*, Third edition, McGraw-Hill Publishing Company, 1979.

[AF] A. Andreotti and T. Frankel, *The Lefschetz theorem on hyperplane sections*, Ann. of Math. (2) **69** (1959), 713–717.

[AEJ] J. K. Arason, R. Elman, and B. Jacob, *The graded Witt ring and Galois cohomology I*, in *Quadratic and Hermetian forms*, Canadian Math. Soc. Conference Proceedings Vol. 4 (1984), 17–50.

[AM] M. Artin and B. Mazur, *Étale homotopy theory*, Lecture Notes in Math. **100**, ??????

[BT] H. Bass and J. Tate, *The Milnor ring of a global field*, Algebraic K-theory, II: "Classical" algebraic K-theory and connections with arithmetic (Proc. Conf., Seattle, Wash., Battelle Memorial Inst., 1972), pp. 349–446. Lecture Notes in Math., Vol. 342, Springer, Berlin, 1973.

[SGA4] M. Artin, A. Grothendieck, and J.L. Verdier, *Théorie des topos and cohomologie étale des schémas*, Lecture Notes in Math. **269, 270, 305**, Springer-Verlag, 1972–1973.

[AH1] M. Atiyah and F. Hirzebruch, *Vector bundles and homogeneous spaces*, Proc. Sympos. Pure Math. AMS **3**, pp. 7–38.

[AH2] M. Atiyah and F. Hirzebruch, *Analytic cycles on complex manifolds*, Topology **1** (1961), 25–45.

[BO] P. Berthelot and A. Ogus, *Notes on crystalline cohomology*, Princeton University Press, Princeton, NJ, 1978.

[B] R. Bott, *On a theorem of Lefschetz*, Michigan Math. J.**6** (1959), 211–216.

[D1] P. Deligne, *La conjecture de Weil pour des surfaces $K3$*, Invent. Math. **15** (1972), 206–226.

[D2] P. Deligne, *Les intersetions complète de niveau de Hodge un*, Invent. Math. **15** (1972), 237–250.

[D3] P. Deligne, *La conjecture de Weil. I.* Inst. Hautes Études Sci. Publ. Math. **43** (1974), 273–307.

[D4] P. Deligne, *La conjecture de Weil. II.* Inst. Hautes Études Sci. Publ. Math. **52** (1980), 137–252.

[SGA7b] P. Deligne and N. Katz, *Groupes de monodromie en géométrie algébrique. II.* Séminaire de Géometrié Algébrique du Bois-Marie, 1967–1969. Lecture Notes in Math. **340**, Springer-Verlag, Berlin-New York, 1973.

[De] A. Delzant, *Définition des classes de Stiefel-Whitney d'un module quadratique sur un corps de charactéristique différent de 2*, C.R. Acad. Sci. Paris **255**, 1366–1368.

[Du] D. Dugger, *An Atiyah-Hirzebruch spectral sequence for KR-theory*, to appear in K-theory.

[DI] D. Dugger and D. C. Isaksen, *Topological hypercovers and $\mathbb{A}^1$-realizations*, Math. Z. **246** (2004), 667–689.

[Dw] B. Dwork, *On the rationality of the zeta function of an algebraic variety*, Amer. J. Math. **82** (1960), 631–648.

[EL] R. Elman and T.Y. Lam, *Pfister forms and K-theory of fields*, Jour. of Algebra **23** (1972), 181–213.

[FV] E. Friedlander and V. Voevodsky, *Bivariant cycle cohomology*. Cycles, transfers, and motivic homology theories, 138–187, Ann. of Math. Stud., 143, Princeton Univ. Press, Princeton, NJ, 2000.

[GR] H. Grauert and R. Remmert, *Komplexe Räume*, Math. Ann.**136** (1958), 245–318.

[GH] M. Greenberg and J. Harper, *Algebraic topology*, Addison-Wesley Publishing, 1981.

[G1] A. Grothendieck, *Sur une note de Mattuck-Tate*, J. Reine u. Angew. Math. **200** (1958), 208–215.

[G2] A. Grothendieck, *Standard conjectures on algebraic cycles*, Algebraic Geometry (Internat. Colloq., Tata Inst. Fund. Res., Bombay, 1968), pp. 193–199, Oxford Univ. Press, London, 1969.

[SGA7a] A. Grothendieck, *Groupes de monodromie en géométrie algébrique. I.* Séminaire de Géometrié Algébrique du Bois-Marie, 1967–1969. Lecture Notes in Math. **288**, Springer-Verlag, Berlin-New York, 1972.

[Ha] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer-Verlag New York, 1977.

[H] F. Hirzebruch, *Topological methods in algebraic geometry*, Springer, New York, 1966.

[Ho] W.V.D. Hodge, *The topological invariants of algebraic varieties*, Proceedings of the International Congress of Mathematicians, Vol. 1, Harvard (1950), 182–191.

[IR] K. Ireland and M. Rosen, *A classical introduction to modern number theory, second edition*. Graduate Texts in Mathematics, Springer-Verlag New York, 1990.

[K1] K. Kato, *A generalization of local class field theory by using K-groups II*, J. Fac. Sci. Univ. Tokyo Sect. 1A Math. **27** (1980), no. 3, 603–683.

[Ka] N. Katz, *An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields*, Mathematical developments arising from Hilbert problems (Proc. Sympos. Pure Math., Vol. XXVIII, Northern Illinois Univ., De Kalb, Ill., 1974), pp. 275–305.

[KM] N. Katz and W. Messing, *Some consequences of the Riemann hypothesis for varieties over finite fields*, Invent. Math. **23** (1974), 73–77.

[Kl1] S. Kleiman, *Algebraic cycles and the Weil conjectures*, Dix esposés sur la cohomologie des schémas, pp. 359–386. North-Holland, Amsterdam; Masson, Paris, 1968.

[Kl2] S. Kleiman, *The standard conjectures*, Motives (Seattle, WA, 1991), 3–20, Proc. Sympos. Pure Math., 55, Part 1, Amer. Math. Soc., Providence, RI, 1994.

[Kl3] S. Kleiman, *Motives*, Algebraic geometry, Oslo 1970 (Proc. Fifth Nordic Summer-School in Math., Oslo, 1970), pp. 53–82. Wolters-Noordhoff, Groningen, 1972.

[La] K. Lamotke, *The topology of complex projective varieties after S. Lefschetz*, Topology **20** (1981), no. 1, 15–51.

[LZ] J. Lannes and S. Zarati, *Invariants de Hopf d'ordre supérieur et suite spectrale d'Adams*, C.R. Acad. Sc. Paris **296** (1983), 695–698.

[L] S. Lefschetz, *L'analysis situs et la géométrie algébrique*, Gauthier-Villars, Paris, 1924.

[Lo] D. Lorenzini, *An invitation to arithmetic geometry*, Graduate Studies in Math. Vol. **9**, American Mathematical Society, Providence, 1996.

[Ma] Y. I. Manin, *Correspondences, motifs and monoidal transformations* (Russian), Mat. Sb. (N.S.) **77** (119), 1968, 475–507. English translation in Mathematics U.S.S.R. Sbornik.

[M1] B. Mazur, *Frobenius and the Hodge filtration*, Bull. Amer. Math. Soc. **78**, no. 5 (1972), 653–667.

[M2] B. Mazur, *Frobenius and the Hodge filtration (estimates)*, Ann. of Math. (2) **98** (1973), 58–95.

[M3] B. Mazur, *Eigenvalues of Frobenius acting on algebraic varieties over finite fields*, Algebraic geometry (Proc. Sympos. Pure Math., Vol. 29, Humboldt State Univ., Arcata, Calif., 1974), pp. 231–261. Amer. Math. Soc., Providence, R.I., 1975

[MVW] C. Mazza, V. Voevodsky, and C. Weibel, *Lectures on motivic cohomology*, preprint, July 2002. http://www.math.uiuc.edu/K-theory/0486.

[M] A. Merkujev, *On the norm residue symbol of degree* 2, Soviet Math. Doklady **24** (1981), 546–551.

[Me] W. Messing, *Short sketch of Deligne's proof of the hard Lefschetz theorem*, Algebraic geometry (Proc. Sympos. Pure Math., Vol. 29, Humboldt State Univ., Arcata, Calif., 1974), pp. 563–580. Amer. Math. Soc., Providence, R.I., 1975.

[Mn] J. Milne, *Étale cohomology*, Princeton Mathematical Series, 33. Princeton University Press, Princeton, N.J., 1980.

[Mr1] J. Milnor. *Morse theory*, Annal of Mathematics Studies, No. 51, Princeton University Press, Princeton, NJ, 1963.

[Mr2] J. Milnor, *Algebraic K-theory and quadratic forms*, Invent. Math. **9** (1970), 318–344.

[M1] F. Morel, *Voevodsky's proof of Milnor's conjecture*, Bull. Amer. Math. Soc. **35**, no. 2 (1998), 123–143.

[M2] F. Morel, *Suite spectral d'Adams et invariants cohomologiques des formes quadratiques*, C.R. Acad. Sci. Ser. 1 Math. **328** (1999), no. 11, 963–968.

[M3] F. Morel, *An introduction to $\mathbb{A}^1$-homotopy theory*, Trieste lectures. Preprint, 2002. Available at http://www.math.jussieu.fr/∼morel/.

[M4] F. Morel, *Milnor's conjecture on quadratic forms and mod* 2 *motivic complexes*, preprint, 2004. http://www.math.uiuc.edu/K-theory/0684.

[M5] F. Morel, *Suites spectrales d'Adams et conjectures de Milnor*. Draft available online at http://www.math.jussieu.fr/∼morel/.

[MT] R. E. Mosher and M. C. Tangora, *Cohomology operations and applications in homotopy*, ????

[Mu] J. Murre, *Algebraic cycles and algebraic aspects of cohomology and K-theory*, Algebraic cycles and Hodge theory (Torino, 1993), 93–152, Lecture Notes in Math., 1594, Springer, Berlin, 1994.

[OVV] D. Orlov, A. Vishik, and V. Voevodsky, *An exact sequence for $K_*^M/2$ with applications to quadratic forms*, preprint, 2000. http://www.math.uiuc.edu/K-theory/0454.

[Pf1] A. Pfister, *Some remarks on the historical development of the algebraic theory of quadratic forms*, in *Quadratic and Hermetian forms*, Canadian Math. Soc. Conference Proceedings Vol. 4 (1984), 1–16.

[Pf2] A. Pfister, *On the Milnor conjectures: history, influence, applications*, Jarhes. Deutsch. Math.-Verein. **102** (2000), 15–41.

[R1] M. Rost, *Some new results on the Chow groups of quadrics*, preprint, 1990. Available at http://www.math.uiuc.edu/K-theory/0165/.

[R2] M. Rost, *Norm varieties and algebraic cobordism*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), 77–85, Higher Ed. Press, Beijing, 2002.

[S1] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der mathematischen Wissenschaften **270**, Springer-Verlag Berlin Heidelberg, 1985.

[S2] W. Scharlau, *On the history of the algebraic theory of quadratic forms*, in *Quadratic forms and their applications*, Contemp. Math. **272**, American Mathematical Society, 2000, 229–259.

[S1] G. Segal, *Classifying spaces and spectral sequences*, Inst. Hautes Etudes Sci. Publ. Math. **34**, 102–112 (1968).

[Se1] J.P. Serre, *Sur la topologie des variétés algébriques en caractéristique p*, 1958 Symposium internacional de topologia algebraica, Universidad Nacional Autónoma de México and UNESCO, Mexico City, 24–53.

[Se4] J.P. Serre, *Analogues Kähleriennes des certaines conjectures de Weil*, Ann. of Math. (2) **71** (1960), 392–394.

[Se2] J.P. Serre, *Zeta and L functions*, Arithmetical Algebraic Geometry (Proc. Conf. Purdue Univ., 1963), pp. 82–92, Harper & Row, New York, 1965.

[Se3] J.P. Serre, *Cohomologie Galoisienne*, Lecture Notes in Mathematics **5**, Springer-Verlag, 1973.

[Se5] J.P. Serre, *Géométrie algébrique et géométrie analytique*, Ann. Inst. Fourier, Grenoble **6** (1955–1956), 1–42.

[Si] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer-Verlag New York, 1986.

[Su1] D. Sullivan, *Geometric topology: Localization, periodicity, and Galois symmetry. The 1970 MIT notes.* Edited and with a preface by Andrew Ranicki. K-Monographs in Mathematics, **8**. Springer, Dordrecht, 2005.

[Su2] D. Sullivan, *Genetics of homotopy theory and the Adams conjecture*, Ann. of Math. (2) **100** (1974), 1–79.

[Su] A. Suslin, *Voevodsky's proof of the Milnor conjecture*, Current Developments in Mathematics, 1997 (Cambridge, MA), 173–188.

[V1] V. Voevodsky, *The Milnor conjecture*, preprint, 1996. http://www.math.uiuc.edu/K-theory/0170.

[V2] V. Voevodsky, *Reduced power operations in motivic cohomology*, Publ. Math. Inst. Hautes Études Sci., No. 98 (2003), 1-57.

[V3] V. Voevodsky, *Motivic cohomology with $\mathbb{Z}/2$-coefficients*, Publ. Math. Inst. Hautes Études Sci., No. 98 (2003), 59–104.

[V4] V. Voevodsky, *On motivic cohomology with $\mathbb{Z}/l$ coefficients*, preprint, 2003. Available at http://www.math.uiuc.eud/K-theory/0639.

[VSF] V. Voevodsky, A. Suslin, and E. M. Friedlander, *Cycles, transfers, and motivic homology theories*, Annals of Mathematics Studies **143**, Princeton University Press, Princeton, NJ, 2000.

[V5] V. Voevodsky, *Cohomological theory of presheaves with transfers*. Cycles, transfers, and motivic homology theories, 87–137, Ann. of Math. Stud., 143, Princeton Univ. Press, Princeton, NJ, 2000.

[V6] V. Voevodsky, *Triangulated categories of motives over a field*. Cycles, transfers, and motivic homology theories, 188–238, Ann. of Math. Stud., 143, Princeton Univ. Press, Princeton, NJ, 2000.

[W1] A. Weil, *Sur les fonctions algébriques á corps de constantes fini*, C. R. Acad. Sci. Paris 210 (1940), 592–594.

[W2] A. Weil, *On the Riemann hypothesis in function fields*, Proc. Nat. Acad. Sci. U. S. A. **27** (1941), 345–347.

[W3] A. Weil, *Variétés abéliennes et courbes algébriques*, Actualités Sci. Ind., no. 1064 = Publ. Inst. Math. Univ. Strasbourg **8** (1946). Hermann & Cie., Paris, 1948.

[W4] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Actualités Sci. Ind., no. 1041 = Publ. Inst. Math. Univ. Strasbourg **7** (1945). Hermann & Cie., Paris, 1948.

[W5] A. Weil, *Numbers of solutions of equations in finite fields*, Bull. Amer. Math. Soc. **55** (1949), 497–508.

[W6] A. Weil, *Abstract versus classical algebraic geometry*, Proceedings of the International Congress of Mathematicians, 1954, Amsterdam, vol. III, pp. 550–558. Erven P. Noordhoff N.V., Groningen; North-Holland Publishing Co., Amsterdam, 1956.