INTRODUCTION TO ARITHMETIC GEOMETRY (NOTES FROM 18.782, FALL 2009)

BJORN POONEN

Contents

Date: December 10, 2009.

1. What is arithmetic geometry?

Algebraic geometry studies the set of solutions of a multivariable polynomial equation (or a system of such equations), usually over $\mathbb{R}$ or $\mathbb{C}$. For instance, $x^2 + xy - 5y^2 = 1$ defines a hyperbola. It uses both commutative algebra (the theory of commutative rings) and geometric intuition.

Arithmetic geometry is the same except that one is interested instead in the solutions where the coordinates lie in other fields that are usually far from being algebraically closed. Fields of special interest are $\mathbb{Q}$ (the field of rational numbers) and $\mathbb{F}_p$ (the finite field of p elements), and their finite extensions. Also of interest are solutions with coordinates in $\mathbb{Z}$ (the ring of integers).

Example 1.1. The circle $x^2 + y^2 = 1$ has infinitely many rational points, such as $(3/5, 4/5)$. Finding them all is essentially the same as finding all Pythagorean triples.

Example 1.2. The circle $x^2 + y^2 = 3$ has no rational points at all!

Example 1.3. The curve $x^4 + y^4 = 1$ has exactly four rational points, namely $(\pm 1, 0)$ and $(0, \pm 1)$. This is the exponent 4 case of Fermat's Last Theorem: this case was proved by Fermat himself.

We'll develop methods for explaining things like this.

2. Absolute values on fields

One approach to constructing the field $\mathbb{Q}_p$ of p-adic numbers is to copy the construction of $\mathbb{R}$, but with a twist: the usual absolute value is replaced by an exotic measure of size.

Definition 2.1. An absolute value on a field k is a function

$$k \to \mathbb{R}_{\geq 0}$$
$$x \mapsto \|x\|$$

such that the following hold for $x, y \in k$:

(Abs1) $\|x\| = 0$ if and only if $x = 0$
(Abs2) $\|xy\| = \|x\| \cdot \|y\|$

(Abs3) $\|x+y\| \le \|x\| + \|y\|$ ("triangle inequality")

Examples:

- $\mathbb{R}$ with the usual $|\ |$
- $\mathbb{C}$ with the usual $|\ |$ (or any subfield of this)
- any field k with

$$\|x\| := \begin{cases} 1, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0. \end{cases}$$

This is called the trivial absolute value.

Definition 2.2. An absolute value $\|\ \|$ satisfying

(Abs3′) $\|x+y\| \le \max(\|x\|, \|y\|)$ ("nonarchimedean triangle inequality")

is said to be nonarchimedean. Otherwise it is said to be archimedean.

(Abs3′) is more restrictive than (Abs3), since $\max(\|x\|, \|y\|) \le \|x\| + \|y\|$.

(Abs3′) is strange from the point of view of classical analysis: it says that if you add many copies of a "small" number, you will never get a "large" number, no matter how many copies you use. This is what gives p-adic analysis its strange flavor.

Of the absolute values considered so far, only the trivial absolute value is nonarchimedean. But we will construct others soon. In fact, most absolute values are nonarchimedean!

3. The p-adic absolute value on $\mathbb{Q}$

The fundamental theorem of arithmetic (for integers) implies that every nonzero rational number x can be factored as

$$x = u \prod_p p^{n_p} = u 2^{n_2} 3^{n_3} 5^{n_5} \cdots$$

where $u \in \{1, -1\}$, and $n_p \in \mathbb{Z}$ for each prime p, and $n_p = 0$ for almost all p (so that all but finitely many factors in the product are 1, making it a finite product).

Definition 3.1. Fix a prime p. The p-adic valuation is the function

$$\begin{aligned} v_p \colon \mathbb{Q}^\times &\to \mathbb{Z} \\ x &\mapsto v_p(x) := n_p, \end{aligned}$$

that gives the exponent of p in the factorization of a nonzero rational number x. If $x = 0$, then by convention, $v_p(0) := +\infty$. Sometimes the function is called ord_p instead of v_p.

Another way of saying the definition: If x is a nonzero rational number, it can be written in the form $p^n \frac{r}{s}$, where r and s are integers *not* divisible by p, and $n \in \mathbb{Z}$, and then $v_p(x) := n$.

Example 3.2. We have $v_2(5/24) = -3$, since $5/24 = 2^{-3}\frac{5}{3} = 2^{-3}3^{-1}5^1$.

Properties:

(Val1) $v_p(x) = +\infty$ if and only if $x = 0$

(Val2) $v_p(xy) = v_p(x) + v_p(y)$

(Val3) $v_p(x + y) \geq \min(v_p(x), v_p(y))$

These hold even when x or y is 0, as long as one uses reasonable conventions for $+\infty$, namely:

- $(+\infty) + a = +\infty$
- $+\infty \geq a$
- $\min(+\infty, a) = a$

for any a, including $a = +\infty$.

Property (Val2) says that if we disregard the input 0, then v_p is a homomorphism from the multiplicative group $\mathbb{Q}^\times$ to the additive group $\mathbb{Z}$.

Proof of (Val3). The cases where $x = 0$ or $y = 0$ or $x + y = 0$ are easy, so assume that x, y, and $x + y$ are all nonzero. Write

$$x = p^n \frac{r}{s} \qquad \text{(and)} \qquad y = p^m \frac{u}{v}$$

with r, s, u, v not divisible by p, so $v_p(x) = n$ and $v_p(y) = m$. Without loss of generality, assume that $n \leq m$. Then

$$\begin{aligned} x + y &= p^n \left(\frac{r}{s} + \frac{p^{m-n} u}{v} \right) \\ &= p^n \frac{N}{sv}. \end{aligned}$$

Here sv is not divisible by p, but N might be so N might contribute some extra factors of p. Thus all we can say is that

$$v_p(x + y) \geq n = \min(n, m) = \min(v_p(x), v_p(y)). \qquad \square$$

Definition 3.3. Fix a prime p. The p-adic absolute value of a rational number x is defined by

$$|x|_p := p^{-v_p(x)}.$$

If $x = 0$ (i.e., $v_p(x) = +\infty$), then we interpret this as $|0|_p := 0$.

Properties (Val1), (Val2), (Val3) for v_p are equivalent to properties (Abs1), (Abs2), (Abs3′) for $|\ |_p$. In particular, $|\ |_p$ really is an absolute value on $\mathbb{Q}$.

4. Ostrowski's classification of absolute values on $\mathbb{Q}$

On $\mathbb{Q}$ we now have absolute values $|\ |_2$, $|\ |_3$, $|\ |_5$, ..., and the usual absolute value $|\ |$, which is also denoted $|\ |_\infty$, for reasons having to do with an analogy with function fields that we will not discuss now. Ostrowski's theorem says that these are essentially all of them.

Definition 4.1. Two absolute values $\| \ \|$ and $\| \ \|'$ on a field k are said to be equivalent if there is a positive real number α such that

$$\|x\|' = \|x\|^\alpha$$

for all $x \in k$.

Theorem 4.2 (Ostrowski)**.** *Every nontrivial absolute value on $\mathbb{Q}$ is equivalent to $| \ |_p$ for some $p \leq \infty$.*

Proof. Let $\| \ \|$ be the absolute value.

Case 1: there exists a positive integer b with $\|b\| > 1$. Let b be the smallest such positive integer. Since $\|1\| = 1$, it must be that $b > 1$. Let α be the positive real number such that $\|b\| = b^\alpha$. Any other positive integer n can be written in base b:

$$n = a_0 + a_1 b + \cdots + a_s b^s$$

where $0 \leq a_i < b$ for all i, and $a_s \neq 0$. Then

$$\begin{aligned}
\|n\| &\leq \|a_0\| + \|a_1 b\| + \|a_2 b^2\| + \cdots + \|a_s b^s\| \\
&= \|a_0\| + \|a_1\| b^\alpha + \|a_2\| b^{2\alpha} + \cdots + \|a_s\| b^{s\alpha} \\
&\leq 1 + b^\alpha + b^{2\alpha} + \cdots + b^{s\alpha} \qquad \text{(by definition of } b\text{, since } 0 \leq a_i < b\text{)} \\
&= \left(1 + b^{-\alpha} + b^{-2\alpha} + \cdots + b^{-s\alpha}\right) b^{s\alpha} \\
&\leq C n^\alpha \qquad \text{(since } b^s \leq n\text{)},
\end{aligned}$$

where C is the value of the convergent *infinite* geometric series

$$1 + b^{-\alpha} + b^{-2\alpha} + \cdots .$$

This holds for all n, so for any $N \geq 1$ we can substitute n^N in place of n to obtain

$$\|n^N\| \leq C (n^N)^\alpha,$$

which implies

$$\begin{aligned}
\|n\|^N &\leq C (n^\alpha)^N \\
\|n\| &\leq C^{1/N} n^\alpha.
\end{aligned}$$

This holds for all $N \geq 1$, and $C^{1/N} \to 1$ as $N \to \infty$, so we obtain

$$\|n\| \leq n^\alpha$$

for each $n \geq 1$.

We next prove the opposite inequality $\|n\| \geq n^\alpha$ for all positive integers n. Given n, choose an integer s such that $b^s \leq n < b^{s+1}$. Then

$$\|b^{s+1}\| \leq \|n\| + \|b^{s+1} - n\|$$

so

$$\begin{aligned}
\|n\| &\geq \|b^{s+1}\| - \|b^{s+1} - n\| \\
&= b^{(s+1)\alpha} - \|b^{s+1} - n\| \qquad (\text{since } \|b\| = b^\alpha) \\
&\geq b^{(s+1)\alpha} - (b^{s+1} - n)^\alpha \qquad (\text{by the previous paragraph}) \\
&\geq b^{(s+1)\alpha} - (b^{s+1} - b^s)^\alpha \qquad (\text{since } b^s \leq n < b^{s+1}) \\
&= b^{(s+1)\alpha}\left[1 - \left(1 - \frac{1}{b}\right)^\alpha\right] \\
&= (bn)^\alpha\left[1 - \left(1 - \frac{1}{b}\right)^\alpha\right] \\
&= cn^\alpha,
\end{aligned}$$

where c is a positive real number independent of n. This inequality, $\|n\| \geq cn^\alpha$ holds for all positive integers n, so as before, we may substitute $n = n^N$, take N^{th} roots, and take the limit as $N \to \infty$ to deduce

$$\|n\| \geq n^\alpha.$$

Combining the previous two paragraphs yields $\|n\| = n^\alpha$ for any positive integer n. If m is another positive integer, then

$$\begin{aligned}
\|n\| \cdot \|m/n\| &= \|m\| \\
\|m/n\| &= \|m\|/\|n\| = m^\alpha/n^\alpha = (m/n)^\alpha.
\end{aligned}$$

Thus $\|q\| = q^\alpha$ for every positive rational number. Finally, if q is a positive rational number, then

$$\|-q\| = \|-1\| \cdot \|q\| = q^\alpha = |-q|^\alpha$$

so $\|x\| = |x|^\alpha$ holds for all $x \in \mathbb{Q}$ (including 0).

Case 2: $\|b\| = 1$ *for all positive integers* b. Then as in the previous paragraph, the axioms of absolute values imply that $\|x\| = 1$ for all $x \in \mathbb{Q}^\times$, contradicting the assumption that $\|\ \|$ is a nontrivial absolute value.

Case 3: $\|n\| \leq 1$ *for all positive integers* n*, and there exists a positive integer* b *such that* $\|b\| < 1$. Assume that b is the smallest such integer. If it were possible to write $b = rs$ for some smaller positive integers r and s, then $\|r| = 1$ and $\|s\| = 1$ by definition of b, but then $\|b\| = \|r\| \cdot \|s\| = 1$, a contradiction; thus b is a prime p.

We prove (by contradiction) that p is the only prime satisfying $\|p\| < 1$. Suppose that q were another such prime. For any positive integer N, the integers p^N and q^N are relatively prime, so there exist integers u, v such that

$$up^N + vq^N = 1,$$

and then

$$\begin{aligned} 1 = \|1\| &= \|up^N + vq^N\| \\ &\leq \|u\| \cdot \|p\|^N + \|v\| \cdot \|q\|^N \\ &\leq \|p\|^N + \|q\|^N. \end{aligned}$$

This is a contradiction if N is large enough. So $\|q\| = 1$ for every prime $q \neq p$.

Since $0 < \|p\| < 1$ and $0 < |p|_p < 1$, there exists a positive real number α such that $\|p\| = |p|_p^\alpha$. Now, for any nonzero rational number

$$x = \pm \prod_{\text{primes } q \text{ including } p} q^{n_q}$$

property (Abs2) (and $\|-1\| = 1$) imply

$$\|x\| = \prod_{\text{primes } q \text{ including } p} \|q\|^{n_q} = \|p\|^{n_p}$$

since all the other factors are 1. Since $\|p\| = |p|_p^\alpha$, this becomes

$$\|x\| = |p|_p^{n_p \alpha} = |x|_p^\alpha.$$

□

5. Cauchy sequences and completion

Let k be a field equipped with an absolute value $\| \ \|$.

Definition 5.1. A sequence (a_i) in k converges if there exists $\ell \in k$ such that for every $\epsilon > 0$, the terms a_i are eventually within ϵ of ℓ: i.e., for every $\epsilon > 0$, there exists a positive integer N such that for all $i \geq N$, the distance bound $\|a_i - \ell\| < \epsilon$ holds. In this case, ℓ is called the limit of the sequence.

Equivalently (a_i) converges to ℓ if and only if $\|a_i - \ell\| \to 0$ as $i \to \infty$. The limit is unique if it exists: if (a_i) converges to both ℓ and ℓ', then

$$\|\ell' - \ell\| \leq \|a_i - \ell\| + \|a_i - \ell\| \to 0 + 0 = 0,$$

so $\|\ell' - \ell\| = 0$, so $\ell' = \ell$.

Definition 5.2. A sequence (a_i) in k is a Cauchy sequence if for every $\epsilon > 0$, the terms are eventually within ϵ of each other; i.e., for every $\epsilon > 0$, there exists a positive integer N such that for all $i, j \geq N$, the distance bound $\|a_i - a_j\| < \epsilon$ holds.

Proposition 5.3. *If a sequence converges, it is a Cauchy sequence.*

Proof. Use the triangle inequality. □

Unfortunately, the converse can fail.

Definition 5.4. A field k is complete with respect to $\| \ \|$ if every Cauchy sequence converges.

We would like every Cauchy sequence to converge, but this might not be the case. To fix this, for each Cauchy sequence that does not converge, we could formally create a new symbol that represents the limit and treat it as if it were a new number. But some Cauchy sequences look as if they should be converging to the same limit, so we need to identify some of these symbols. So the new symbols really should correspond to equivalence classes of Cauchy sequences that do not converge. Actually there is no harm in creating symbols for Cauchy sequences that converge already, as long as these new symbols are identified with the pre-existing limits. Finally, we can think of the equivalence classes themselves as being the symbols.

Definition 5.5. Two sequences (a_i) and (b_i) are equivalent if $\|a_i - b_i\| \to 0$ as $i \to \infty$.

One can check that this induces an equivalence relation on the set of sequences. Any sequence equivalent to a Cauchy sequence is also a Cauchy sequence.

Definition 5.6. The completion $\hat{k}$ of k with respect to $\| \ \|$ is defined to be the set of equivalence classes of Cauchy sequences in k.

One can define all the field operations on $\hat{k}$. For instance, the product of the equivalence classes of the Cauchy sequences (a_i) and (b_i) is the equivalence class of $(a_i b_i)$. (One can check that this is a Cauchy sequence, and that its equivalence class is unchanged if (a_i) and (b_i) are replaced by equivalent Cauchy sequences.) The 1 in $\hat{k}$ is the equivalence class of the sequence $1, 1, 1, \ldots$. If (x_i) is a Cauchy sequence not equivalent to $(0, 0, 0, \ldots)$, then the x_i are eventually nonzero, and setting $y_i := \begin{cases} x_i^{-1} & \text{if } x_i \neq 0 \\ 0 & \text{if } x_i = 0 \end{cases}$ defines a Cauchy sequence whose equivalence class is an inverse of the equivalence class of (x_i).

Moreover, the operations satisfy all the field axioms, so $\hat{k}$ is a new field. The map $k \to \hat{k}$ sending a to the equivalence class of the constant sequence $(a, a, \ldots)$ is a ring homomorphism, and ring homomorphisms between fields are always injective, so k is identified with a subfield of $\hat{k}$.

Define an absolute value $\| \ \|'$ on $\hat{k}$ by decreeing that the absolute value of the equivalence class of (a_i) is $\lim_{i\to\infty} \|a_i\|$. The restriction of $\|;\|'$ to the embedded copy of k is just the original absolute value $\|;\|$. If $\alpha \in \hat{k}$ is represented by the Cauchy sequence (a_i) in k, then the sequence (a_i) viewed in $\hat{k}$ converges to α.

The absolute value $\| \ \|'$ is nonarchimedean if and only if $\| \ \|$ was. (One way to see that is by using the characterization that $\|;\|$ is nonarchimedean if and only if $\|n\| \leq 1$ for all positive integers n.)

Finally, $\hat{k}$ is complete with respect to $\| \ \|$.

Example 5.7. The completion of $\mathbb{Q}$ with respect to the usual absolute value $|\ |$ is the field $\mathbb{R}$ of real numbers.

Proposition 5.8. *Let k be a subfield of a complete field L. Then*

(1) The inclusion $k \hookrightarrow L$ extends to an embedding $\hat{k} \hookrightarrow L$.
(2) If every element of L is a limit of a sequence in k, then the embedding $\hat{k} \hookrightarrow L$ is an isomorphism.

Proof. (1) Given an element $a \in \hat{k}$, represented as the limit of (a_i) with $a_i \in k$, map a to the limit of (a_i) in L. This defines a ring homomorphism $\hat{k} \to L$, which is automatically injective since these are fields.

(2) Suppose that every element of L is a limit of a sequence in k. Given $\ell \in L$, choose a sequence (a_i) in k converging to ℓ. Then (a_i) is Cauchy, so it also converges to an element $a \in \hat{k}$. This a maps to ℓ, by definition of the embedding. So the embedding is surjective as well as injective; hence it is an isomorphism.

□

6. Inverse limits

Definition 6.1. An inverse system of sets is an infinite sequence of sets (A_n) with maps between them as follows:

$$\cdots \to A_{n+1} \xrightarrow{f_n} A_n \to \cdots \xrightarrow{f_1} A_1 \xrightarrow{f_0} A_0.$$

Definition 6.2. The inverse limit $A = \varprojlim A_n$ of an inverse system of sets $(A_n), (f_n)$ as above is the set A whose elements are the infinite sequences (a_n) with $a_n \in A_n$ for each $n \geq 0$ satisfying the compatibility condition $f_n(a_{n+1}) = a_n$ for each $n \geq 0$. It comes with a projection map $\epsilon_n \colon A \to A_n$ that takes the n^{th} term in the sequence.

Remark 6.3. If the A_n are groups and the f_n are group homomorphisms, then the inverse limit A has the structure of a group: multiply sequences term-by-term. If the A_n are rings and the f_n are ring homomorphisms, then the inverse limit A has the structure of a ring.

7. Defining $\mathbb{Z}_p$ as an inverse limit

Fix a prime p. Let A_n be the ring $\mathbb{Z}/p^n\mathbb{Z}$. Let f_n be the ring homomorphism sending $\bar{b} := b + p^{n+1}\mathbb{Z}$ to $\bar{b} := b + p^n\mathbb{Z}$. The ring of p-adic integers is $\mathbb{Z}_p := \varprojlim A_n$.

For example, if $p = 3$, then a sequence like

$$0 \bmod 1,\ 2 \bmod 3,\ 5 \bmod 9,\ 23 \bmod 27,\ \cdots$$

defines an element of $\mathbb{Z}_3$.

8. Properties of $\mathbb{Z}_p$

Recall that a sequence of group homomorphism is exact if at the group in each position, the kernel of the outgoing arrow equals the image of the incoming arrow. For example,

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is called a short exact sequence if f is injective, g is surjective, and g induces an isomorphism from B/A (or more precisely, $B/f(A)$) to C.

Proposition 8.1. *For each $m \geq 0$,*

$$0 \to \mathbb{Z}_p \xrightarrow{p^m} \mathbb{Z}_p \xrightarrow{\epsilon_m} \mathbb{Z}/p^m\mathbb{Z} \to 0$$

is exact. (Here the first map is the multiplication-by-p^m map, sending $(a_n)_{n\geq 0}$ to $(p^m a_n)_{n\geq 0}$., and ϵ_m maps $(a_n)_{n\geq 0}$ to a_m.)

Proof. First let us check that multiplication-by-p on $\mathbb{Z}_p$ is injective. Suppose that $a = (a_n) \in \mathbb{Z}_p$ is in the kernel. Then $pa = 0$, so $pa_n = 0$ in $\mathbb{Z}/p^n\mathbb{Z}$ for all n. In particular, $pa_{n+1} = 0$ in $\mathbb{Z}/p^{n+1}\mathbb{Z}$. That means that $a_{n+1} = p^n y_{n+1}$ for some $y_{n+1} \in \mathbb{Z}/p^{n+1}\mathbb{Z}$. But then $a_n = f_n(a_{n+1}) = p^n f_n(y_{n+1}) = 0$ in $\mathbb{Z}/p^n\mathbb{Z}$. This holds for all n, so $a = 0$.

Exactness on the left: Since multiplication-by-p is injective, composing this with itself m times shows that multiplication-by-p^m is injective.

Exactness on the right: Given an element $\beta \in \mathbb{Z}/p^m\mathbb{Z}$, choose an integer b that represents β. Then the constant sequence b represents an element of $\mathbb{Z}_p$ mapping to β.

Exactness in the middle: If $a \in \mathbb{Z}_p$, then $\epsilon_m(p^m a) = p^m \epsilon(a) = 0$ in $\mathbb{Z}/p^m\mathbb{Z}$. Thus the image of the incoming arrow (multiplication-by-p^m) is contained in the kernel of the outgoing arrow (ϵ_m).

Conversely, suppose that $x = (x_n)$ is in the kernel of ϵ_m. So $x_m = 0$. Then for all $n \geq m$, we have $x_n \in \frac{p^m\mathbb{Z}}{p^n\mathbb{Z}}$. So there is a unique y_{n-m} mapping to x_n via the isomorphism

$$\frac{\mathbb{Z}}{p^{n-m}\mathbb{Z}} \xrightarrow{p^m} \frac{p^m\mathbb{Z}}{p^n\mathbb{Z}}.$$

These y_{n-m} are compatible (because the x_n are), so as n ranges through integers $\geq m$, they form an element $y \in \mathbb{Z}_p$ such that $p^m y = x$. So x is in the image of multiplication-by-p^m. □

Proposition 8.2.

(1) *An element of $\mathbb{Z}_p$ is a unit if and only if it is not divisible by p. In other words, the group of p-adic units $\mathbb{Z}_p^\times$ equals $\mathbb{Z}_p - p\mathbb{Z}_p$.*
(2) *Every nonzero $a \in \mathbb{Z}_p$ can be uniquely expressed as $p^n u$ with $n \in \mathbb{Z}_{\geq 0}$ and $u \in \mathbb{Z}_p^\times$.*

Proof.

(1) If $a = (a_n) \in \mathbb{Z}_p$ is divisible by p, then $a_1 = 0$, so a cannot have an inverse. Conversely, if $a = (a_n)$ is not divisible by p, then $a_n \in \mathbb{Z}/p^n\mathbb{Z}$ is represented by an integer not divisible by p, so a_n has an inverse $b_n \in \mathbb{Z}/p^n\mathbb{Z}$. These b_n must be compatible, and $b := (b_n)$ is an inverse of a in $\mathbb{Z}_p$.

(2) *Existence:* If $a = (a_n) \in \mathbb{Z}_p$ is nonzero, then there is a largest n such that $a_n = 0$. For that n, Proposition 8.1 implies that $a = p^n u$ for some $u \in \mathbb{Z}_p$. Moreover, u cannot be divisible by p (since otherwise $a_{n+1} = 0$ too), so u is a unit.

Uniqueness: Suppose that $p^n u = p^m u'$. If $m = n$, then using injectivity of multiplication-by-p^m we get $u = u'$, so the factorizations are the same. Otherwise, without loss of generality $n > m$. Then $u' = p^{n-m}u$ is a unit divisible by p, contradicting (1).

□

Multiplying nonzero elements $p^n u$ and $p^m u'$ yields $p^{n+m}uu'$, whose $(n+m+1)^{\text{th}}$ component is nonzero, so $\mathbb{Z}_p$ is an integral domain. In fact, $\mathbb{Z}_p$ is a UFD with one prime!

9. The field of p-adic numbers

Definition 9.1. The field $\mathbb{Q}_p$ of p-adic numbers is the fraction field of $\mathbb{Z}_p$.

Each nonzero $a \in \mathbb{Q}_p$ is uniquely expressible as $p^n u$ with $n \in \mathbb{Z}$ and $u \in \mathbb{Z}_p^\times$. (For existence, any nonzero $a \in \mathbb{Q}_p$ is $(p^{m'}u')/(p^m u)$ for some $m, m' \in \mathbb{Z}_{\geq 0}$ and $u, u' \in \mathbb{Z}_p^\times$, so $a = p^{m'-m}(u'u^{-1})$.

Define the p-adic valuation on $\mathbb{Q}_p$ by $v_p(p^n u) = n$ whenever $n \in \mathbb{Z}$ and $u \in \mathbb{Z}_p^\times$, and $v_p(0) := +\infty$. Then define $|x|_p := p^{-v_p(x)}$ for each $x \in \mathbb{Q}_p$.

The ring $\mathbb{Z}$ injects into $\mathbb{Z}_p$, so its fraction field $\mathbb{Q}$ injects into $\mathbb{Q}_p$, and the p-adic valuation and absolute value on $\mathbb{Q}_p$ restrict to the p-adic valuation and absolute value on $\mathbb{Q}$ previously defined.

Proposition 9.2.

(1) The field $\mathbb{Q}_p$ is complete with respect to $|\ |_p$.

(2) Every element of $\mathbb{Q}_p$ is a limit of a sequence in $\mathbb{Q}$.

Proof.

(1) Let (a_n) be a Cauchy sequence in $\mathbb{Q}_p$. Then (a_n) is bounded. By multiplying by a suitable power of p, we can reduce to the case where $a_n \in \mathbb{Z}_p$ for all n. Choose an infinite subsequence S_1 whose image in $\mathbb{Z}/p\mathbb{Z}$ is constant. Choose an infinite subsequence S_2 of S_1 whose image in $\mathbb{Z}/p^2\mathbb{Z}$ is constant, and so on. Form a sequence by choosing one element from S_1, a later element from S_2, and so on. Then this subsequence converges in $\mathbb{Z}_p$ to the element whose image in each $\mathbb{Z}/p^n\mathbb{Z}$ is the image

of the subsequence S_n. Finally, a Cauchy sequence with a convergent subsequence converges.

(2) Let $a \in \mathbb{Q}_p$. By multiplying by a suitable power of p, we reduce to the case where $a \in \mathbb{Z}_p$. Write $a = (a_n)$ with $a_n \in \mathbb{Z}/p^n\mathbb{Z}$. Choose an integer $b_n \in \mathbb{Z}$ representing a_n. Then $v_p(a - b_n) \geq n$, so $|a - b_n| \leq p^{-n}$, so the sequence (b_n) converges to a in $\mathbb{Q}_p$.

□

Combining Propositions 5.8 and 9.2 shows that $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ with respect to $|\ |_p$.

10. p-ADIC EXPANSIONS

Definition 10.1. Say that a series $\sum_{n=1}^{\infty} a_n$ of p-adic numbers converges if and only if the sequence of partial sums converges with respect to $|\ |_p$.

Theorem 10.2.

(1) Each $a \in \mathbb{Z}_p$ has a unique expansion $a = b_0 + b_1 p + b_2 p^2 + \cdots$ with $b_n \in \{0, 1, \ldots, p-1\}$ for all n.

(2) Each $a \in \mathbb{Q}_p$ has a unique expansion $a = \sum_{n \in \mathbb{Z}} b_n p^n$ in which $b_n \in \{0, 1, \ldots, p-1\}$ and $b_n = 0$ for all sufficiently negative n.

(3) For either expansion, $v_p(a)$ is the least integer n such that $b_n \neq 0$. (If no such n exists, then $a = 0$ and $v_p(a) = +\infty$.)

Proof.

(1) *Existence:* Write $a = (a_n)$ with $a_n \in \mathbb{Z}/p^n\mathbb{Z}$. Choose $s_n \in \{0, 1, \ldots, p^n - 1\}$ representing a_n. Write $s_n = b_0 + b_1 p + \cdots + b_{n-1} p^{n-1}$ with $b_i \in \{0, 1, \ldots, p-1\}$. The compatibility condition on the a_n implies that the b_i so defined are independent of n; i.e., the base-p expansion of s_{n+1} extends the base-p expansion of s_n by one term $b_n p^n$. Then $s_n \to a$ in $\mathbb{Q}_p$, so

$$b_0 + b_1 p + b_2 p^2 + \cdots = a.$$

Uniqueness: If $b'_n \in \{0, 1, \ldots, p-1\}$ also satisfy

$$b'_0 + b'_1 p + b'_2 p^2 + \cdots = a,$$

then we get

$$b_0 + b_1 p + \cdots + b_{n-1} p^{n-1} \equiv b'_0 + b'_1 p + \cdots + b'_{n-1} p^{n-1} \pmod{p^n},$$

but both sides are integers in $\{0, 1, \ldots, p^n - 1\}$, so they are equal, and this forces $b_i = b'_i$ for all i.

(2) *Existence:* If $a \in \mathbb{Q}_p$, then there exists $m \in \mathbb{Z}$ such that $p^m a \in \mathbb{Z}_p$. Write

$$p^m a = b_0 + b_1 p + b_2 p^2 + \cdots$$

with $b_i \in \{0, 1, \ldots, p-1\}$ and divide by p^m.

Uniqueness: Follows from uniqueness for $\mathbb{Z}_p$.

(3) If $a = b_0 + b_1 p + \cdots \in \mathbb{Z}_p$ with $b_i \in \{0, 1, \ldots, p-1\}$, and $b_0 \neq 0$, then a has nonzero image in $\mathbb{Z}/p\mathbb{Z}$, so a is a unit, and $v_p(a) = 0$. The general case follows from this one by multiplying by p^n for an arbitrary $n \in \mathbb{Z}$.

□

11. Solutions to polynomial equations

Lemma 11.1 ("Compactness argument")**.** *Let $\cdots \to S_2 \to S_1 \to S_0$ be an inverse system of finite nonempty sets. Then $\varprojlim S_i$ is nonempty.*

Proof. Let $T_{i,0}$ be the image of $S_i \to \cdots \to S_0$. Then

$$\cdots \subseteq T_{2,0} \subseteq T_{1,0} \subseteq T_{0,0},$$

but these are finite nonempty sets, so $T_{i,0}$ must be constant for sufficiently large i. Let E_0 be this "eventual image". Define $T_{i,1}$ and E_1 in the same way, and define E_2, and so on. Then the E_i form an inverse system in which the maps $E_{i+1} \to E_i$ are *surjective*. Choose $e_0 \in E_0$, choose a preimage $e_1 \in E_1$ of e_0, choose a preimage $e_2 \in E_2$ of e_1, and so on: this defines an element of $\varprojlim S_i$. □

Proposition 11.2. *Let $f \in \mathbb{Z}_p[x]$ be a polynomial. Then the following are equivalent:*

(1) The equation $f(x) = 0$ has a solution in $\mathbb{Z}_p$.
(2) The equation $f(x) = 0$ has a solution in $\mathbb{Z}/p^n\mathbb{Z}$ for every $n \geq 0$.

Proof. Let S_n be the set of solutions in $\mathbb{Z}/p^n\mathbb{Z}$. Then $\varprojlim S_n \subseteq \varprojlim \mathbb{Z}/p^n\mathbb{Z} = \mathbb{Z}_p$ is the set of solutions in $\mathbb{Z}_p$. We have $\varprojlim S_n \neq \emptyset$ if and only if all the S_n are nonempty, by Lemma 11.1. □

12. Hensel's lemma

Hensel's lemma says that approximate zeros of polynomials can be improved to exact zeros.

Theorem 12.1 (Hensel's lemma)**.** *Let $f \in \mathbb{Z}_p[x]$. Suppose that $f(a) \equiv 0 \pmod{p}$, and $f'(a) \not\equiv 0 \pmod{p}$. (That is, a is a simple root of $(f \bmod p)$.) Then there exists a unique $b \in \mathbb{Z}_p$ with $b \equiv a \pmod{p}$ such that $f(b) = 0$.*

Proof. We prove by induction that for $n \geq 1$ there exists $a_n \in \mathbb{Z}_p$ such that $a_n \equiv a \pmod{p}$ and $f(a_n) \equiv 0 \pmod{p^n}$ (and that $a_n \bmod p^n$ is uniquely determined). For $n = 1$, take $a_1 = a$. Now suppose that the result is known for some $n \geq 1$. So

$$f(a_n) = p^n c,$$

for some $c \in \mathbb{Z}_p$. We try to adjust a_n slightly to make the value of f even smaller p-adically. More precisely, we try $a_{n+1} = a_n + \epsilon$ for a p-adic integer ϵ to be determined: Taylor's theorem gives

$$f(a_{n+1}) = f(a_n) + f'(a_n)\epsilon + g(\epsilon)\epsilon^2$$

for some polynomial $g(x) \in \mathbb{Z}_p[x]$. (This is really just expanding $f(a_n + \epsilon)$ as a polynomial in ϵ.) Choose $\epsilon = p^n z$ with $z \in \mathbb{Z}_p$. Then

$$\begin{aligned} f(a_{n+1}) &= f(a_n) + f'(a_n)p^n z + g(p^n z)p^{2n} z^2 \\ &\equiv p^n c + f'(a_n)p^n z \pmod{p^{n+1}}. \end{aligned}$$

Since

$$f'(a_n) \equiv f'(a) \not\equiv 0 \pmod{p},$$

we get

$$f(a_{n+1}) \equiv (c + f'(a)z)p^n \pmod{p^{n+1}},$$

and there is a unique $z \bmod p$ that makes $c + f'(a)z \equiv 0 \pmod{p}$, and hence a unique choice of $a_{n+1} \bmod p^{n+1}$ that makes $f(a_{n+1}) \equiv 0 \pmod{p^{n+1}}$ This completes the inductive step.

Since $f(x) = 0$ has a unique solution in each $\mathbb{Z}/p^n\mathbb{Z}$ congruent to a modulo p, these solutions give a unique solution in $\mathbb{Z}_p$ congruent to a modulo p. □

This is the p-adic analogue of Newton's method, in which one approximates the polynomial by a linear function in order to pass from an approximate zero to an even better approximation to a zero.

13. Structure of $\mathbb{Q}_p^\times$

The map $\epsilon_n \colon \mathbb{Z}_p \to \mathbb{Z}/p^n\mathbb{Z}$ restricts to a surjective homomorphism

$$\mathbb{Z}_p^\times \to (\mathbb{Z}/p^n\mathbb{Z})^\times.$$

Its kernel is $U_n := 1 + p^n\mathbb{Z}_p$. So $\mathbb{Z}_p^\times/U_n \simeq (\mathbb{Z}/p^n\mathbb{Z})^\times$, and

$$\mathbb{Z}_p^\times \simeq \varprojlim \mathbb{Z}_p^\times/U_n \simeq \varprojlim (\mathbb{Z}/p^n\mathbb{Z})^\times.$$

The U_n form a descending chain of subgroups inside $\mathbb{Z}_p^\times$:

$$\cdots \subset U_3 \subset U_2 \subset U_1 \subset \mathbb{Z}_p^\times.$$

Let $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$. (Generally one writes $\mathbb{F}_p$ when $\mathbb{F}_p$ is being thought of as a field, and $\mathbb{Z}/p\mathbb{Z}$ when it is being thought of as a ring or an abelian group.)

Lemma 13.1. *The quotients in the filtration are:*

(1) $\mathbb{Z}_p^\times / U_1 \simeq \mathbb{F}_p^\times$*, and*
(2) $U_n / U_{n+1} \simeq \mathbb{Z}/p\mathbb{Z}$ *for all* $n \geq 1$*.*

Proof. The first of these has already been proved. For the second, observe that

$$U_n \to \mathbb{Z}/p\mathbb{Z}$$
$$1 + p^n z \mapsto (z \bmod p)$$

is surjective and has kernel U_{n+1}. □

Corollary 13.2. *The order of* U_1/U_n *is* p^{n-1}*.*

Proposition 13.3. *Let* μ_{p-1} *be the set of solutions to* $x^{p-1} = 1$ *in* $\mathbb{Z}_p^\times$*. Then* μ_{p-1} *is a group (under multiplication) mapping isomorphically to* $\mathbb{F}_p^\times$*, and* $\mathbb{Z}_p^\times = U_1 \times \mu_{p-1}$*.*

Proof. The set μ_{p-1} is the kernel of the $(p-1)^{\text{th}}$ power map from $\mathbb{Z}_p^\times$ to itself, so it is a group. Given $a \in \{1, 2, \ldots, p-1\}$, Hensel's lemma shows that μ_{p-1} contains a unique p-adic integer congruent to a modulo p. And there are no elements of μ_{p-1} congruent to 0 mod p. So reduction modulo p induces an isomorphism $\mu_{p-1} \to \mathbb{F}_p^\times$.

We have $U_1 \cap \mu_{p-1} = \{1\}$ (by Hensel's lemma, there is only one solution to $x^{p-1} - 1 = 0$ congruent to 1 modulo p). Also, $U_1 \cdot \mu_{p-1} = \mathbb{Z}_p^\times$, since any $a \in \mathbb{Z}_p^\times$ can be divided by an element of μ_{p-1} congruent to a modulo p to land in U_1. Thus the direct product $U_1 \times \mu_{p-1}$ is equal to $\mathbb{Z}_p^\times$. □

Lemma 13.4. *Let* p *be a prime. If* $p \neq 2$*, let* $n \geq 1$*; if* $p = 2$*, let* $n \geq 2$*. If* $x \in U_n - U_{n+1}$*, then* $x^p = U_{n+1} - U_{n+2}$*.*

Proof. We have $x = 1 + kp^n$ for some k not divisible by p. Then

$$\begin{aligned} x^p &= 1 + \binom{p}{1} kp^n + \binom{p}{2} k^2 p^{2n} + \cdots + k^p p^{np} \\ &\equiv 1 + kp^{n+1} \pmod{p^{n+2}}. \end{aligned}$$

so $x^p \in U_{n+1} - U_{n+2}$. □

Proposition 13.5. *If* $p \neq 2$*, then* $U_1 \simeq \mathbb{Z}_p$*. If* $p = 2$*, then* $U_1 = \{\pm 1\} \times U_2$ *and* $U_2 \simeq \mathbb{Z}_2$*.*

Proof. Suppose then $p \neq 2$. Let $\alpha = 1 + p \in U_1 - U_2$. By the previous lemma applied repeatedly, $\alpha^{p^i} \in U_{i+1} - U_{i+2}$. Let α_n be the image of α in U_1/U_n. Then $\alpha_n^{p^{n-2}} \neq 1$ but $\alpha_n^{p^{n-1}} = 1$, so α_n has exact order p^{n-1}. On the other hand, the group it belongs to, U_1/U_n, also has order p^{n-1}. So U_1/U_n is cyclic, generated by α_n. We have an isomorphism of inverse

systems

$$\begin{array}{ccccccc} \cdots & \longrightarrow & \mathbb{Z}/p^n\mathbb{Z} & \longrightarrow & \mathbb{Z}/p^{n-1}\mathbb{Z} & \longrightarrow & \cdots \\ & & \downarrow & & \downarrow & & \\ \cdots & \longrightarrow & U_1/U_{n+1} & \longrightarrow & U_1/U_n & \longrightarrow & \cdots \end{array}$$

Taking inverse limits shows that $\mathbb{Z}_p \simeq U_1$.

For $p = 2$, the same argument with $\alpha = 1 + 4$ works to prove that $\mathbb{Z}_2 \simeq U_2$. Now $\{\pm 1\}$ and U_2 have trivial intersection, and they generate U_1 (since U_2 has index 2 in U_1), so the direct product $\{\pm 1\} \times U_2$ equals U_1. □

Theorem 13.6.

(1) The group $\mathbb{Z}_p^\times$ is isomorphic to $\mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}_p$ if $p \neq 2$, and to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}_2$ if $p = 2$.

(2) The group $\mathbb{Q}_p^\times$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}_p$ if $p \neq 2$, and to $\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}_2$ if $p = 2$.

Proof.

(1) Combine Propositions 13.3 and 13.5.

(2) The map

$$\begin{aligned} \mathbb{Z} \times \mathbb{Z}_p^\times &\to \mathbb{Q}_p^\times \\ (n, u) &\mapsto p^n u \end{aligned}$$

is an isomorphism of groups. Now substitute the known structure of $\mathbb{Z}_p^\times$ into this.

□

14. Squares in $\mathbb{Q}_p^\times$

14.1. The case of odd p.

Theorem 14.1.

(1) An element $p^n u \in \mathbb{Q}_p^\times$ (with $n \in \mathbb{Z}$ and $u \in \mathbb{Z}_p^\times$) is a square if and only if n is even and $u \bmod p$ is a square in $\mathbb{F}_p^\times$.

(2) We have $\mathbb{Q}_p^\times/\mathbb{Q}_p^{\times 2} \simeq (\mathbb{Z}/2\mathbb{Z})^2$.

(3) For any $c \in \mathbb{Z}_p^\times$ with $c \bmod p \notin \mathbb{F}_p^{\times 2}$, the images of p and c generate $\mathbb{Q}_p^\times/\mathbb{Q}_p^{\times 2}$.

Proof.

(1) We have $\mathbb{Q}_p^\times = p^{\mathbb{Z}} \times \mathbb{F}_p^\times \times \mathbb{Z}_p$, and $2\mathbb{Z}_p = \mathbb{Z}_p$, so

$$\mathbb{Q}_p^{\times 2} = p^{2\mathbb{Z}} \times \mathbb{F}_p^{\times 2} \times \mathbb{Z}_p.$$

Thus an element $p^n u$ is a square if and only if n is even and $u \bmod p \in \mathbb{F}_p^{\times 2}$.

(2) Using the same decomposition, $\mathbb{Q}_p^\times/\mathbb{Q}_p^{\times 2} = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{F}_p^\times/\mathbb{F}_p^{\times 2}) \times \{0\} \simeq (\mathbb{Z}/2\mathbb{Z})^2$ since $\mathbb{F}_p^\times$ is cyclic of even order.

(3) Under the isomorphism above, p and c correspond to the generators of the two copies of $\mathbb{Z}/2\mathbb{Z}$.

□

14.2. **The case $p = 2$.**

Theorem 14.2.

(1) An element $2^n u \in \mathbb{Q}_2^\times$ (with $n \in \mathbb{Z}$ and $u \in \mathbb{Z}_2^\times$) is a square if and only if n is even and $u \equiv 1 \pmod 8$.
(2) We have $\mathbb{Q}_2^\times/\mathbb{Q}_2^{\times 2} \simeq (\mathbb{Z}/2\mathbb{Z})^3$.
(3) The images of 2, -1, 5 generate $\mathbb{Q}_2^\times/\mathbb{Q}_2^{\times 2}$.

Proof.

(1) We have $\mathbb{Q}_2^\times = 2^{\mathbb{Z}} \times \{\pm 1\} \times U_2$, where $U_2 \simeq \mathbb{Z}_2$. Under this last isomorphism, U_3 corresponds to $2\mathbb{Z}_2$, so
$$\mathbb{Q}_2^{\times 2} = 2^{2\mathbb{Z}} \times \{1\} \times U_3.$$
Thus an element $2^n u$ is a square if and only if n is even and $u \equiv 1 \pmod 8$.

(2) Using the same decomposition,
$$\mathbb{Q}_2^\times/\mathbb{Q}_2^{\times 2} = (\mathbb{Z}/2\mathbb{Z}) \times \{\pm 1\} \times \mathbb{Z}_2/2\mathbb{Z}_2 \simeq (\mathbb{Z}/2\mathbb{Z})^3.$$

(3) Under the isomorphism above, 2, -1, and 5 correspond to the generators of the three copies of $\mathbb{Z}/2\mathbb{Z}$.

□

15. p-adic analytic functions

A power series $f(z) := \sum a_n z^n$ with $a_n \in \mathbb{Q}_p$ defines a differentiable function on the open set in $\mathbb{Q}_p$ on which it converges.

Identities between complex power series with rational coefficients can be used to deduce identities between p-adic power series. For example, consider the formal power series
$$z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots$$
$$z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$
in $\mathbb{Q}[[z]]$. Over $\mathbb{C}$, they define the analytic functions $\log(1+z)$ and $e^z - 1$, respectively, in some neighborhoods of 0. They are inverses to each other. So their composition in either order is a function represented by $z \in \mathbb{Q}[[z]]$. On the other hand, their composition in either order is represented also by the formal composition of the power series in $\mathbb{Q}[[z]]$. Two formal power series representing the same analytic function are the same, so the formal power series are inverse to each other. Finally, this identity saying that the composition of the formal

power series in either order gives z implies that the corresponding p-adic analytic functions are inverses of each other when both converge.

16. Algebraic closure

Given a field k, let $k[x]_{\geq 1}$ be the set of polynomials in $k[x]$ of degree at least 1.

Definition 16.1. A field k is algebraically closed if and only if every $f \in k[x]_{\geq 1}$ has a zero in k.

Definition 16.2. An algebraic closure of a field k is a algebraic field extension $\overline{k}$ of k that is algebraically closed.

Example 16.3. The field $\mathbb{C}$ of complex numbers is an algebraic closure of $\mathbb{R}$. But $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$ because some elements of $\mathbb{C}$ (like e and π) are not algebraic over $\mathbb{Q}$.

Theorem 16.4. *Every field k has an algebraic closure, and any two algebraic closures of k are isomorphic over k (but the isomorphism is not necessarily unique).*

Step 1: Given $f \in k[x]_{\geq 1}$, there exists a field extension $E \supseteq k$ in which f has a zero.

Proof. Choose an irreducible factor g of f. Define $E := k[x]/(g(x))$. Then E is a field extension of k, and the image of x in E is a zero of f. □

Step 2: Given $f_1, \ldots, f_n \in k[x]_{\geq 1}$, there exists a field extension $E \supseteq k$ in which each f_i has a zero.

Proof. Step 1 and induction. □

Step 3: There exists a field extension $k' \supseteq k$ containing a zero of every $f \in k[x]_{\geq 1}$.

Proof. Define a commutative ring

$$A := \frac{k[\{X_f : f \in k[x]_{\geq 1}\}]}{(f(X_f) : f \in k[x]_{\geq 1})}.$$

Suppose that A is the zero ring. Then 1 is in the ideal generated by the $f(X_f)$. So we have an equation

$$1 = g_1 f_1(X_{f_1}) + \cdots + g_n f_n(X_{f_n}).$$

for some polynomials g_i. By Step 2, there exists a field extension $F \supseteq k$ containing a zero α_i of each f_i. Evaluating the previous equation at $X_{f_i} = \alpha_i$ yields $1 = 0$ in F, contradicting the fact that F is a field.

Thus A is not the zero ring. So A has a maximal ideal $\mathfrak{m}$. Let $k' := A/\mathfrak{m}$. Then k' is a field extension of k, and the image of X_{f_i} in k' is a zero of F_i. □

Step 4: There exists an algebraically closed field $E \supseteq k$.

Proof. Iterate Step 3 to obtain a chain of fields

$$k \subseteq k' \subseteq k'' \subseteq \cdots \subseteq k^{(n)} \subseteq \cdots .$$

Let E be their union. Any polynomial in $E[x]_{\geq 1}$ has coefficients in some fixed $k^{(n)}$, and hence has a zero in $k^{(n+1)}$, so it has a zero in E. Thus E is algebraically closed. □

Step 5: There exists an algebraic closure $\overline{k}$ of k.

Proof. Let E be as in Step 4. Let $\overline{k}$ be the set of $\alpha \in E$ that are algebraic over k. Since algebraic elements are closed under addition, multiplication, etc., the set $\overline{k}$ is a subfield of E. And of course, $\overline{k}$ is algebraic over k.

If $f \in \overline{k}[x]_{\geq 1}$, let β be a zero of f in E; then β is algebraic over the field $k(\text{coefficients of } f)$, which is algebraic over k, so β is algebraic over k, so $\beta \in \overline{k}$. Thus $\overline{k}$ is algebraically closed. □

Step 6: If E is an algebraic extension of k, and L is an algebraically closed field then any embedding $k \hookrightarrow L$ extends to an embedding $E \hookrightarrow L$.

Proof. If E is generated by one element α, then $E \simeq k[x]/(f(x))$ for some $f \in k[x]_{\geq 1}$. Choose a zero $\alpha' \in L$ of f, and define $E \hookrightarrow L$ by mapping α to α'.

If E is generated by finitely many elements, extend the embedding in stages, adjoining one element at a time.

In general, use transfinite induction (Zorn's lemma). □

Step 7: Any two algebraic closures of k are isomorphic over k.

Proof. Let E and L be two algebraic closures of k. Step 6 extends $k \hookrightarrow L$ to $E \hookrightarrow L$. If $E \neq L$, then the minimal polynomial of an element of $L - E$ would be a polynomial in $E[x]_{\geq 1}$, contradicting the assumption that E is algebraically closed. □

17. Finite fields

Let $\mathbb{F}_p$ be $\mathbb{Z}/p\mathbb{Z}$ viewed as a field.

Theorem 17.1. *For each prime p, choose an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$.*

1. *Given a prime power $q = p^n$, there exists a unique subfield of $\overline{\mathbb{F}}_p$ of order q, namely $\mathbb{F}_q := \{x \in \overline{\mathbb{F}}_p : x^q = x\}$.*
2. *Every finite field is isomorphic to exactly one $\mathbb{F}_q$.*
3. *$\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^n}$ if and only if $m|n$.*
4. *$\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q) \simeq \mathbb{Z}/n\mathbb{Z}$, and it is generated by*

$$\begin{aligned} \mathrm{Frob}_q \colon \mathbb{F}_{q^n} &\to \mathbb{F}_{q^n} \\ x &\mapsto x^q. \end{aligned}$$

Proof.

(1) The p^{th} power map

$$\begin{aligned} \mathrm{Frob}_p \colon \overline{\mathbb{F}}_p &\to \overline{\mathbb{F}}_p \\ x &\mapsto x^p. \end{aligned}$$

is a field homomorphism, by the binomial theorem. In particular, it is injective. Since $\overline{\mathbb{F}}_p$ is algebraically closed, Frob_p is also surjective. So Frob_p is an automorphism of $\overline{\mathbb{F}}_p$. If $q = p^n$, then the q^{th} power map Frob_q is Frob_p^n, so it too is an automorphism of $\overline{\mathbb{F}}_p$. Then $\mathbb{F}_q$ is the subset of $\overline{\mathbb{F}}_p$ fixed by Frob_q, so $\mathbb{F}_q$ is a field. Since $x^q - x$ and $\frac{d}{dx}(x^q - x) = -1$ have no common zeros, the polynomial $x^q - x$ has q distinct zeros in $\overline{\mathbb{F}}_p$. Thus $\#\mathbb{F}_q = q$. This proves the existence half of (1).

(2) (and uniqueness in (1)) Conversely, if K is any finite field, then the characteristic of K is a prime $p > 0$, and the image of $\mathbb{Z} \to K$ is a subfield isomorphic to $\mathbb{F}_p$. Viewing K as an $\mathbb{F}_p$-vector space shows that $\#K = p^n$ for some $n \geq 1$. Let $q = p^n$. The embedding $\mathbb{F}_p \hookrightarrow \overline{\mathbb{F}}_p$ extends to an embedding $K \hookrightarrow \overline{\mathbb{F}}_p$. Since $K^\times$ is a group of order $q - 1$, every element of $K^\times$ satisfies $x^{q-1} = 1$, so every element of K satisfies $x^q = x$, so $K \subseteq \mathbb{F}_q$. But $\#K = \#\mathbb{F}_q = q$, so $K = \mathbb{F}_q$. Finally, K cannot be isomorphic to any $\mathbb{F}_{q'}$ with $q' \neq q$, because its size is q.

(3) If $\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^n}$, then $\mathbb{F}_{p^n}$ is a vector space over $\mathbb{F}_{p^m}$, so p^n is a power of p^m (namely, p^m raised to the dimension), so $m|n$.

Conversely, if $m|n$, write $n = rm$; then

$$\begin{aligned} \mathbb{F}_{p^m} &= \{\text{fixed points of } \mathrm{Frob}_{p^m}\} \\ &\subseteq \{\text{fixed points of } (\mathrm{Frob}_{p^m})^r\} \\ &= \{\text{fixed points of } \mathrm{Frob}_{p^{rm}}\} \\ &= \mathbb{F}_{p^{rm}} \\ &= \mathbb{F}_{p^n}. \end{aligned}$$

(4) The order of $\mathrm{Frob}_q \in \mathrm{Aut}(\mathbb{F}_{q^n})$ is the smallest m such that $x^{q^m} = x$ for all $x \in \mathbb{F}_{q^n}$, which is n. In general, if G is a finite subgroup of $\mathrm{Aut}(K)$, then K is Galois over the fixed field K^G and $\mathrm{Gal}(K/K^G) = G$. Apply this to $K = \mathbb{F}_{q^n}$ and G the cyclic group of order n generated by $\mathrm{Frob}_q \in \mathrm{Aut}(\mathbb{F}_{q^n})$: the fixed field is $\mathbb{F}_q$, so we get

$$\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q) = G \simeq \mathbb{Z}/n\mathbb{Z}.$$

□

The primitive element theorem says that every finite separable extension of a field k is generated by one element α, i.e., is of the form $k[x]/(f(x))$ for some monic irreducible polynomial $f(x) \in k[x]$ (the minimal polynomial of α). So we get

Corollary 17.2. *Given a prime power q and $n \geq 1$, there exists a monic irreducible polynomial $f(x) \in \mathbb{F}_q[x]$ of degree n.*

Remark 17.3. It is not known whether one can *find* such a polynomial in deterministic polynomial time! This is unsolved even for q prime and $n = 2$: i.e., the problem of finding a nonsquare in $\mathbb{F}_p$ in time polynomial in $\log p$ is unsolved.

On the other hand, if one repeatedly chooses a random monic polynomial over $\mathbb{F}_q$ of degree n, then there is a fast test for irreducibility, and one can estimate the probability of irreducibility to show that this succeeds in random polynomial time.

Example 17.4. $\mathbb{F}_2[t]/(t^3 + t + 1)$ is a finite field of order 8.

Warnings: $\mathbb{F}_8 \not\simeq \mathbb{Z}/8\mathbb{Z}$ (the latter is not even a field), and $\mathbb{F}_4 \not\subset \mathbb{F}_8$.

Proposition 17.5. *If k is a field, and G is a finite subgroup of $k^\times$, then G is cyclic.*

Proof. As an abstract group,

$$G \simeq \frac{\mathbb{Z}}{a_1\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{a_n\mathbb{Z}}$$

for some positive integers a_i satisfying $a_1 > 1$ and $a_i | a_{i+1}$ for all i. If $n > 1$, then G has more than a_1 elements of order dividing a_1. But $x^{a_1} - 1$ can have at most $\deg(x^{a_1} - 1) = a_1$ zeros in k. Thus $n = 1$, so G is cyclic. □

Remark 17.6. There is an alternative proof that avoids the structure theorem for finite abelian groups, and instead uses a more elementary counting argument to prove that if G is a finite group of order n such that for each $d|n$, the group G has at most d elements satisfying $x^d = 1$, then G is cyclic.

Corollary 17.7. *The group $\mathbb{F}_q^\times$ is cyclic of order $q - 1$.*

18. Inverse limits in general

Earlier we defined the inverse limit $\varprojlim S_i$ of a sequence of sets S_i indexed by the natural numbers equipped with maps $S_{i+1} \to S_i$. Now we will define $\varprojlim S_i$ given a collection of sets $(S_i)_{i \in I}$ for more general index sets, equipped with maps.

Definition 18.1. A partially ordered set (poset) is a set I equipped with a binary relation $\leq$ such that for all $x, y, z \in I$,

(PO1) $x \leq x$ (reflexivity)
(PO2) If $x \leq y$ and $y \leq x$, then $x = y$ (antisymmetry)
(PO3) If $x \leq y$ and $y \leq z$, then $x = z$ (transitivity).

Definition 18.2. A directed poset is a nonempty poset I such that

(PO4) For every $x, y \in I$, there exists z with $x \geq z$ and $y \geq z$ (any finite subset has an upper bound).

Example 18.3. The set $\mathbb{Z}_{\geq 0}$ with the usual ordering is a directed poset.

Example 18.4. The set $\mathbb{Z}_{>0}$ with the ordering in which $m \leq n$ means $m|n$ is a directed poset.

Definition 18.5. An inverse system of sets is a collection of sets $(S_i)_{i \in I}$ indexed by a directed poset I, together with maps $\phi_i^j \colon S_j \to S_i$ for each $i \leq j$ satisfying the following for all $i, j, k \in I$:

(IS1) $\phi_i^i \colon S_i \to S_i$ is the identity

(IS2) For any $i \leq j \leq k$, the composition $S_k \xrightarrow{\phi_j^k} S_j \xrightarrow{\phi_i^j} S_i$ equals ϕ_i^k.

Definition 18.6. The inverse limit of an inverse system of sets $(S_i)_{i \in I}$ with maps $(\phi_i^j)_{i \leq j}$ is

$$\varprojlim_{i \in I} S_i := \left\{ (s_i) \in \prod_{i \in I} S_i : \phi_i^j(a_j) = a_i \text{ for all } i \leq j \right\},$$

equipped with the projection map π_i to each S_i.

So to give an element of $\varprojlim_{i \in I} S_i$ is to give an element of each S_i such that the elements are compatible with respect to the maps in the inverse system.

Example 18.7. If I is $\mathbb{Z}_{\geq 0}$ with the usual ordering, then this definition of inverse limit reduces to the earlier one. It might seem that there are more maps ϕ_i^j, but they are determined as compositions of the maps ϕ_i^{i+1}.

If the S_i are groups and the ϕ_i^j are group homomorphisms, then $\varprojlim_{i \in I} S_i$ is a group.

Example 18.8. Let I be $\mathbb{Z}_{>0}$ ordered by divisibility. For each $n \in I$, let $G_n = \mathbb{Z}/n\mathbb{Z}$. For $n|N$, define

$$\phi_n^N \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$$
$$\bar{a} \mapsto \bar{a}.$$

Then the inverse limit $\varprojlim_n \mathbb{Z}/n\mathbb{Z}$ is a group called $\hat{\mathbb{Z}}$.

Example 18.9. Let G be any group. Let I be the collection of normal subgroups $N \lhd G$ of finite index. Define $N \leq N'$ if $N' \subseteq N$. This makes I into a directed poset. (Given $N_1, N_2 \in I$, the intersection $N_1 \cap N_2$ is an "upper bound" for N_1 and N_2.) If $N' \subseteq N$, we have a surjective homomorphism

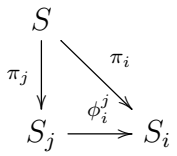
$$G/N' \to G/N$$
$$\bar{g} \mapsto \bar{g}.$$

The inverse limit $\hat{G} := \varprojlim_N G/N$ is a group called the profinite completion of G.

Example 18.10. The profinite completion of $\mathbb{Z}$ is $\hat{\mathbb{Z}}$.

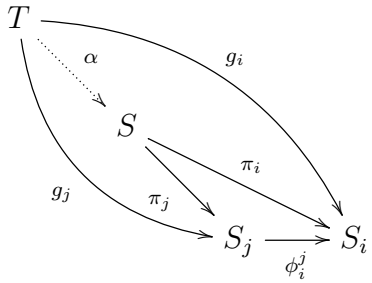
Theorem 18.11. *Let S be the inverse limit of an inverse system of sets $(S_i)_{i\in I}$ with maps $(\phi_i^j)_{i\le j}$. Then S has the following universal property:*

(1) There are maps $\pi_i\colon S \to S_i$ for $i \in I$, compatible with respect to the ϕ_i^j; i.e.,



commutes for all $i \le j$.

(2) For any other set T equipped with maps $g_i\colon T \to S_i$ for $i \in I$, compatible with respect to the ϕ_i^j, there exists a unique map $\alpha\colon T \to S$ such that $g_i(t) = \pi_i(\alpha(t))$ for all $i \in I$ and $t \in T$:



Example 18.12. Let $\hat{G}$ be the profinite completion of a group G. Then G has a natural quotient map to G/N for each finite-index normal subgroup $N \lhd G$. These maps are compatible with the maps $G/N' \to G/N$ of the inverse system, so the universal property yields a homomorphism $G \to \hat{G}$.

Proposition 18.13. *There is an isomorphism $\hat{\mathbb{Z}} \to \prod_{\text{primes } p} \mathbb{Z}_p$.*

Proof. Fix a prime p. Let I be $\mathbb{Z}_{>0}$ ordered by divisibility. For $n \in I$, let $G_n := \mathbb{Z}/p^{v_p(n)}\mathbb{Z}$. For $m|n$, let $G_n \to G_m$ be the quotient map sending 1 to 1. To give a compatible collection of elements of the G_n is equivalent to giving a compatible collection of elements of $\mathbb{Z}/p^m\mathbb{Z}$ for $m \ge 0$, so $\varprojlim_{n\in I} G_n \simeq \mathbb{Z}_p$.

For each $n \in I$, the Chinese remainder theorem gives a natural isomorphism

$$\frac{\mathbb{Z}}{n\mathbb{Z}} \simeq \prod_{\text{primes } p} \frac{\mathbb{Z}}{p^{v_p(n)}\mathbb{Z}}.$$

Taking the inverse limit of both sides yields

$$\hat{\mathbb{Z}} \simeq \prod_{\text{primes } p} \mathbb{Z}_p.$$

□

19. Profinite groups

Definition 19.1. A profinite group is an inverse limit $\varprojlim_{i\in I} G_i$ of *finite* groups G_i.

Examples:

(1) $\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}$ for any prime p
(2) $\hat{\mathbb{Z}} = \varprojlim \mathbb{Z}/n\mathbb{Z}$
(3) $\mathrm{GL}_r(\mathbb{Z}_p) = \varprojlim_n \mathrm{GL}_r(\mathbb{Z}/p^n\mathbb{Z})$ for any fixed prime p and fixed $r \geq 0$.
(4) The profinite completion of any group.

19.1. Order.

Definition 19.2. Assuming that the inverse system maps $G_j \to G_i$ are all surjective, the order $\#G$ of a profinite group $G := \varprojlim_{i\in I} G_i$ is the least common multiple of $\#G_i$, interpreted as a supernatural number $\prod_p p^{e_p}$ where each e_p is either a nonnegative integer or ∞.

Example 19.3. $\#\mathbb{Z}_5^\times = 2^2 5^\infty$.

19.2. Topology on a profinite group.

(This subsection is for those who know the basic definitions of topology.) The profinite topology on a profinite group $G = \varprojlim_{i\in I} G_i$ is constructed as follows. Equip each finite group G_i with the discrete topology. Equip $\prod_{i\in I} G_i$ with the product topology. Then $G = \varprojlim_{i\in I} G_i$ is a closed subset of $\prod_{i\in I} G_i$, and we give it the subspace topology. By Tychonoff's theorem, $\prod_{i\in I} G_i$ is compact, so its closed subset G is compact too.

19.3. Subgroups.

The profinite group G is equipped with group homomorphisms $\pi_i \colon G \to G_i$. If H_i is a subgroup of G_i, then $\pi_i^{-1}(H_i)$ is a subgroup of G. These are called the open subgroups of G.

If for every i we choose a subgroup H_i of G_i such that each $\phi_i^j \colon G_j \to G_i$ maps H_j into H_i, then $\varprojlim_{i\in I} H_i$ is a subgroup of $G = \varprojlim_{i\in I} G_i$. These are called the closed subgroups of G.

The open subgroups are exactly the closed subgroups of finite index. In particular, every open subgroup is a closed subgroup, but not vice versa in general.

Example 19.4. The profinite topology on $\mathbb{Z}_p := \varprojlim \mathbb{Z}/p^n\mathbb{Z}$ agrees with the topology coming from $|\ |_p$. The open subgroups of $\mathbb{Z}_p$ are the subgroups $p^e\mathbb{Z}_p$ for $e = 0, 1, 2, \ldots$. The closed subgroups are these together with the trivial subgroup $\{0\}$.

Subgroups of a profinite group that are not even closed are generally worthless! When one encounters such a subgroup, one takes its closure right away.

20. Review of field theory

We recall some definitions of field theory. Let L/k be an algebraic field extension.

Definition 20.1. The extension L/k is normal if it satisfies one of the following equivalent conditions:

(1) Every irreducible polynomial in $k[x]$ with a zero in L factors completely into linear factors in $L[x]$.
(2) If we embed L in an algebraic closure of k, so $k \subseteq L \subseteq \overline{k}$, then every $\sigma \in \mathrm{Aut}(\overline{k}/k)$ satisfies $\sigma(L) = L$.

Definition 20.2. A polynomial $f(x) \in k[x]$ is separable if it satisfies one of the following equivalent conditions:

(1) When factored in $\overline{k}[x]$ for an algebraic closure $\overline{k}$ of k, it has no repeated factors.
(2) The polynomial $f(x)$ and its derivative $f'(x)$ have no common zeros in $\overline{k}$.
(3) We have $\gcd(f(x), f'(x)) = 1$ in $k[x]$.

We will usually be applying the notion of separable to minimal polynomials, which are irreducible. Over a field k of characteristic 0, every irreducible polynomial is separable. *Proof:* We have $\deg f'(x) < \deg f(x)$, and $\mathrm{char}\, k = 0$ implies $f'(x) \neq 0$, so $f'(x)$ is not divisible by $f(x)$. so $\gcd(f(x), f'(x)) = 1$.

Thus separability is an issue mainly in the case of characteristic $p > 0$.

Definition 20.3. An element α in L is separable over k if it satisfies one of the following equivalent conditions:

(1) It is a zero of a separable polynomial in $k[x]$.
(2) The minimal polynomial of α over k is separable.
(3) Either $\mathrm{char}\, k = 0$, or $\mathrm{char}\, k = p$ and the minimal polynomial of α over k is not of the form $g(x^p)$ for a polynomial $g(x) \in k[x]$.

The set of elements of L that are separable over k form an intermediate subfield.

Definition 20.4. If every element of L is separable over k, then L is called separable over k.

By the remark preceding the definition, it is enough if L is generated by separable elements.

If k is a field of characteristic p, the image of the p-power Frobenius endomorphism $k \to k$ is a subfield $k^p := \{a^p : a \in k\}$ of k.

Definition 20.5. A field k is perfect if it satisfies one of the following equivalent conditions:

- Either $\mathrm{char}\, k = 0$, or $\mathrm{char}\, k = p$ and $k = k^p$.
- Every finite extension of k is separable over k.
- Every algebraic extension of k is separable over k.

Example 20.6. Finite fields are perfect.

Example 20.7. The prototypical example of an imperfect field is $k = \mathbb{F}_p(t)$. The prototypical example of an inseparable extension is the extension $L = k(t^{1/p})$ of this k. The minimal polynomial of $t^{1/p}$ over k is $x^p - t$, which is irreducible (as minimal polynomials always are), but not separable.

Definition 20.8. Call L/k Galois if it is both normal and separable. In this case, the Galois group $\mathrm{Gal}(L/k)$ is the set of automorphisms σ of L such that $\sigma(x) = x$ for all $x \in k$.

Definition 20.9. If blah is a property of a group (e.g., abelian), call L/k blah if L/k is a Galois extension and $\mathrm{Gal}(L/k)$ is blah.

Definition 20.10. Let k be a field. Choose an algebraic closure $\overline{k}$. The separable closure of k (in a fixed algebraic closure $\overline{k}$) is $k^{\mathrm{sep}} := \{\alpha \in \overline{k} : \alpha \text{ is separable over } k\}$. It is the maximal subfield of $\overline{k}$ that is separable over k.

The extension k^{sep}/k is Galois.

Definition 20.11. The absolute Galois group of k is $G_k := \mathrm{Gal}(k^{\mathrm{sep}}/k)$.

21. Infinite Galois theory

Let K/k be a Galois extension (possibly of infinite degree). Let I be the set of fields F such that $k \subset F \subset K$ and F is a finite Galois extension of k. Order I by inclusion.

Proposition 21.1.

(1) If $F, F' \in I$, then their compositum FF' (the subfield of K generated by F and F') is in I too.
(2) I is a directed poset
(3) If $k \subset E \subset K$ and E is finite over k, then $E \subseteq F$ for some $F \in I$.
(4) $\bigcup_{F \in I} F = K$.

Proof.

(1) This is a well-known fact about Galois extensions.
(2) This follows from (1).
(3) The primitive element theorem expresses as E as $k[x]/(f(x))$. Let F be the splitting field of $f(x)$.
(4) This follows from (3).

□

For each $F \in I$, the group $\mathrm{Gal}(F/k)$ is finite. If $F \subset F'$, then we have

$$\phi_F^{F'} : \mathrm{Gal}(F'/k) \twoheadrightarrow \mathrm{Gal}(F/k)$$
$$\sigma \mapsto \sigma|_F.$$

Proposition 21.2. *For any Galois extension K/k, there is an isomorphism*

$$\begin{aligned} \mathrm{Gal}(K/k) &\to \varprojlim_{F \in I} \mathrm{Gal}(F/k) \\ \sigma &\mapsto (\sigma|_F)_{F \in I}. \end{aligned}$$

Proof. Each F is normal over k, so $\sigma|_F$ maps F to F. The $\sigma|_F$ are compatible. So by the universal property of the inverse limit, we have a well-defined homomorphism

$$\mathrm{Gal}(K/k) \to \varprojlim_{F \in I} \mathrm{Gal}(F/k).$$

Conversely, compatible elements of $\mathrm{Gal}(F/k)$ for all $F \in I$, glue to give a unique automorphism in $\mathrm{Gal}(K/k)$. □

Corollary 21.3. *Any Galois group $\mathrm{Gal}(K/k)$ can be viewed as a profinite group. In this setting, the profinite topology is also called the Krull topology.*

Theorem 21.4 (Main theorem of Galois theory)**.** *Let K/k be a Galois extension. Let $G = \mathrm{Gal}(K/k)$. Then there exists an inclusion-reversing bijection*

$$\begin{aligned} \{\text{fields } E \text{ such that } k \subseteq E \subseteq K\} &\leftrightarrow \{\textit{closed} \text{ subgroups of } G\} \\ E &\mapsto \mathrm{Gal}(K/E) \\ K^H &\hookleftarrow H. \end{aligned}$$

Moreover, if $E \leftrightarrow H$, then

$$\begin{aligned} E/k \text{ is normal} &\iff H \text{ is normal in } G \\ &\qquad (\text{and then } \mathrm{Gal}(E/K) \simeq G/H) \\ E/k \text{ is finite} &\iff H \text{ is open in } G. \end{aligned}$$

21.1. **Examples of Galois groups.** If $k = \mathbb{C}$, then $G_k = \mathrm{Gal}(\mathbb{C}/\mathbb{C}) = \{1\}$.

If $k = \mathbb{R}$, then $G_k = \mathrm{Gal}(\mathbb{C}/\mathbb{R}) \simeq \mathbb{Z}/2\mathbb{Z}$, generated by complex conjugation.

If $k = \mathbb{F}_q$, then

$$G_k = \mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q) \simeq \varprojlim \mathbb{Z}/n\mathbb{Z} = \hat{\mathbb{Z}} \simeq \prod_{\text{prime } p} \mathbb{Z}_p,$$

a pro-cyclic group (an inverse limit of finite cyclic groups). It is topologically generated by Frob_q; i.e., G_k is the closure of the infinite cyclic subgroup generated by Frob_q.

If $k = \mathbb{Q}_p$, then it turns out that G_k is a pro-solvable group (an inverse limit of finite solvable groups), whose structure is known exactly but is rather complicated. Also, for each $n \geq 1$, there are only finitely many degree-n extensions of $\mathbb{Q}_p$ in $\overline{\mathbb{Q}}_p$.

If $k = \mathbb{Q}$, then G_k is incredibly complicated. Conjecturally every finite group is a quotient of it; i.e., every finite group is $\mathrm{Gal}(F/\mathbb{Q})$ for some finite Galois extension F of $\mathbb{Q}$.

Let $\mathbb{Q}^{\mathrm{ab}}$ be the subfield of $\overline{\mathbb{Q}}$ generated by all finite abelian extensions $F/\mathbb{Q}$. Then $\mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q})$ is abelian; in fact, it is the largest abelian quotient of $G_{\mathbb{Q}}$ (where we allow quotients only by closed subgroups).

Let ζ_n be a primitive n^{th} root of 1 in $\overline{\mathbb{Q}}$. "Irreducibility of the cyclotomic polynomial" implies that $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \simeq (\mathbb{Z}/n\mathbb{Z})^\times$.

Theorem 21.5 (Kronecker-Weber)**.** $\mathbb{Q}^{\mathrm{ab}} = \bigcup_{n\geq 1} \mathbb{Q}(\zeta_n)$.

For instance, $\mathbb{Q}(\sqrt{7})$ is an abelian extension of $\mathbb{Q}$, so the Kronecker-Weber theorem implies that $\sqrt{7}$ must be an element of $\mathbb{Q}(\zeta_n)$ for some n. (In fact, the smallest such n is 28.)

Corollary 21.6.

$$\mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) \simeq \varprojlim_n \left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^\times \simeq \hat{\mathbb{Z}}^\times \simeq \prod_p \mathbb{Z}_p^\times.$$

Class field theory generalizes this to describe the maximal abelian extension k^{ab} of any number field k.

22. Affine varieties

From now on, k is a perfect field, and $\overline{k}$ is a fixed algebraic closure. Let $G_k := \mathrm{Gal}(\overline{k}/k)$.

22.1. **Affine space.**

Definition 22.1. Fix $n \in \mathbb{Z}_{\geq 0}$. For each field extension L of k, define

$$\mathbb{A}^n_k(L) := L^n$$

Here $\mathbb{A}^n_k$ is called n-dimensional affine space over k. (If k is understood, we just write $\mathbb{A}^n$.)

Think of $k[x_1, \ldots, x_n]$ as being the ring of functions on $\mathbb{A}^n_k$. This relationship is written

$$\mathbb{A}^n_k = \mathrm{Spec}\, k[x_1, \ldots, x_n].$$

Remark 22.2. The group G_k acts on $\mathbb{A}^n(\overline{k})$, and the set of fixed points $\mathbb{A}^n(\overline{k})^{G_k}$ is $\mathbb{A}^n(k)$.

22.2. **Affine varieties.**

Loosely speaking, an affine variety is the set of common zeros of a set of polynomials.

Given a subset T of $k[x_1, \ldots, x_n]$, define $Z = Z_T$ by the rule

$$Z(L) := \{P \in L^n : f(P) = 0 \text{ for all } f \in T\}.$$

Any such Z is called an affine variety over k. (Some authors also require an "irreducibility" condition.)

Definition 22.3. An element of $Z(L)$ is called an L-rational point on Z, or simply an L-point.

Example 22.4. Take $k = \mathbb{R}$, $n = 2$, and $T = \{x^2 + y^2 - 1\}$. Then $Z(\mathbb{R})$ is the unit circle in $\mathbb{R}^2$. We say "Z is the variety defined by $x^2 + y^2 = 1$ over $\mathbb{R}$".

The set of polynomials in $k[x_1, \ldots, x_n]$ that vanish at a point P is closed under addition and closed under multiplication by an arbitrary polynomial. So if I is the ideal of $k[x_1, \ldots, x_n]$ generated by T, then $Z_I = Z_T$.

Example 22.5. The zero set of $x^2 + y^2 - 1$ and the zero set of $(x^2 + y^2 - 1)^2$ in L^2 for any field extension L of k are equal.

More generally, any ideal I defines the same set of zeros as its radical

$$\sqrt{I} := \{f \in k[x_1, \ldots, x_n] : f^m \in I \text{ for some } m \geq 0\}.$$

So we will assume that I is radical ($I = \sqrt{I}$) from now on.

Theorem 22.6 (version of Hilbert Nullstellensatz)**.** *There is an inclusion-reversing bijection*

$$\{\text{radical ideals of } k[x_1, \ldots, x_n]\} \leftrightarrow \{\text{affine varieties } Z \text{ in } \mathbb{A}^n_k\}$$
$$I \mapsto Z_I \text{ (where } Z_I(L) = \{\text{common zeros of } f \in I\}$$
$$\{f \in k[x_1, \ldots, x_n] : f(P) = 0 \textit{ for all } P \in Z(L) \textit{ for all } L\} \hookleftarrow Z.$$

We can view elements of $k[x_1, \ldots, x_n]$ as functions on Z, but the functions in I are identically 0 on Z, so the ring of functions on Z is actually $k[x_1, \ldots, x_n]/I$. Thus we write

$$Z = \operatorname{Spec} \frac{k[x_1, \ldots, x_n]}{I}.$$

The commutative ring $\frac{k[x_1,\ldots,x_n]}{I}$ is called the affine coordinate ring of Z.

Example 22.7. Let $X = \operatorname{Spec} \frac{\mathbb{R}[x,y]}{(x^2+y^2+1)}$ and let $Y = \operatorname{Spec} \frac{\mathbb{R}[x,y]}{(1)}$. Is $X = Y$?

No! One reason: $X(\mathbb{C})$ is nonempty, but $Y(\mathbb{C})$ is empty. Another reason: the ideal $(x^2 + y^2 + 1)$ is not the unit ideal (1), since $x^2 + y^2 + 1$ has no inverse in $\mathbb{R}[x, y]$.

Moral: When k is not algebraically closed, it is important to consider $Z(L)$ for all finite extensions L of k instead of just viewing of Z as the set of zeros with coordinates in k.

Remark 22.8. If Z is any affine k-variety, then $Z(k) = Z(\overline{k})^{G_k}$.

22.3. **Irreducible varieties.** The variety defined by $xy = 0$ is the union of the two varieties defined by $x = 0$ and $y = 0$ in $\mathbb{A}^2$.

Definition 22.9. An irreducible variety is a nonempty variety that cannot be decomposed as a union of two smaller varieties.

One can show that a general variety Z is a finite union of irreducible subvarieties, none contained in any other: these are called the irreducible components of Z.

One can show:

Proposition 22.10. *Suppose that $Z = \operatorname{Spec} k[x_1, \dots, x_n]/I$, where I is radical. Then the following are equivalent:*

- *Z is irreducible.*
- *I is a prime ideal.*
- *$k[x_1, \dots, x_n]/I$ is an integral domain.*

If Z is irreducible, the function field $\kappa(Z)$ of Z is defined as the fraction field $\operatorname{Frac} k[x_1, \dots, x_n]/I$.

Example 22.11. The function field of $\mathbb{A}^n_k$ is the rational function field

$$k(x_1, \dots, x_n) := \left\{ \frac{f}{g} : f, g \in k[x_1, \dots, x_n] \right\}.$$

22.4. **Dimension.** There are a couple of equivalent ways to define dimension of a variety X.

Definition 22.12. The dimension $\dim X$ of X is the largest integer d such that there exists a chain of (closed) irreducible varieties

$$Z_0 \subsetneq Z_1 \subsetneq \cdots \subsetneq Z_d$$

contained in X. (If $X = \emptyset$, then $\dim X = -\infty$.)

An alternative, equivalent definition:

Definition 22.13. Let X be an irreducible variety. Then $\dim X$ is the smallest integer d such that the function field $\kappa(X)$ contains elements $f_1, \dots, f_d$ such that $\kappa(X)$ is algebraic over the subfield $k(f_1, \dots, f_n)$ generated by k and the f_i inside $\kappa(X)$.

Then, for any variety X, define $\dim X$ as the maximum of the dimensions of its irreducible components.

(Proving the equivalence requires a lot of commutative algebra.)

Example 22.14. We have $\dim \mathbb{A}^n = n$. A maximal chain of irreducible subvarieties is

$$\mathbb{A}^0 \subseteq \mathbb{A}^1 \subseteq \cdots \subseteq \mathbb{A}^{n-1} \subseteq \mathbb{A}^n,$$

corresponding to the chain of prime ideals

$$(x_1, \dots, x_n) \supseteq (x_2, \dots, x_n) \supseteq \cdots \supseteq (x_n) \supseteq (),$$

of $k[x_1, \dots, x_n]$. (It takes some work to show that there is no longer chain.)

Alternatively, the function field $k(x_1, \dots, x_n)$ is algebraic over the subfield generated by n elements $x_1, \dots, x_n$. (It takes some work to show that one cannot do it with less than n elements.)

22.5. **Smooth varieties.**

Definition 22.15. A hypersurface in $\mathbb{A}_k^n$ is a subvariety defined by a single equation $f(x_1, \ldots, x_n) = 0$ with f a nonzero polynomial in $k[x_1, \ldots, x_n]$.

Definition 22.16. Let X be a hypersurface $f(x_1, \ldots, x_n) = 0$ in $\mathbb{A}_k^n$. A point $P \in X(L)$ (for some field extension L of k is a singularity of X if $\frac{\partial f}{\partial x_i}(P) = 0$ for all i.

The set of singularities forms a subvariety of X, defined by $f = 0$ together with the equations $\frac{\partial f}{\partial x_i} = 0$ for $i = 1, \ldots, n$.

Definition 22.17. A hypersurface X in $\mathbb{A}_k^n$ is called smooth (of dimension $n-1$) or non-singular if there are no singularities in $X(L)$ for any $L \supseteq k$ (actually it suffices to check $X(\overline{k})$).

Example 22.18. Let X be the curve $y^2 = x^3 + 1$ in $\mathbb{A}_{\mathbb{Q}}^2$. Is X singular? Let $f(x, y) := y^2 - x^3 - 1$. The singular locus is defined by the equations

$$\begin{aligned} y^2 - x^3 - 1 &= 0 \\ -3x^2 &= 0 \\ 2y &= 0, \end{aligned}$$

which have no common solutions in $\overline{\mathbb{Q}}$, so the curve is smooth.

(But it would not have been so if instead of $\mathbb{Q}$ we were working over the field $\mathbb{F}_2$ or $\mathbb{F}_3$.)

Example 22.19. Let Y be the "nodal cubic" $y^2 = x^3 + x^2$. The singular locus is defined by the equations

$$\begin{aligned} y^2 - x^3 - x^2 &= 0 \\ -3x^2 - 2x &= 0 \\ 2y &= 0, \end{aligned}$$

which have the common solution $(0, 0)$. So Y is singular, with a unique singularity at $(0, 0)$. Near $(0, 0)$, the curve Y looks approximately like $y^2 = x^2$ (obtained by discarding higher order terms like x^3) so it has two "branches" crossing at $(0, 0)$. Such a singularity is called a node.

More generally:

Definition 22.20. A variety $X := \operatorname{Spec} \frac{k[x_1, \ldots, x_n]}{(f_1, \ldots, f_m)}$ is smooth (of dimension $n-m$) if and only if at every point $P \in X(L)$ for every extension L of k, the Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}\right) \in M_{m \times n}(L)$ has rank m. (Again it suffices to check $P \in X(\overline{k})$.)

The condition that one of the $m \times m$ minors be nonvanishing is exactly the condition in the implicit function theorem to guarantee that X is the graph of a differentiable function, if we were working over $\mathbb{R}$ or $\mathbb{C}$.

Remark 22.21. One can show that if X is smooth of dimension r, then $\dim X = r$.

23. Projective varieties

23.1. **Motivation.** Let $O = (0,0,0) \in \mathbb{R}^3$.

There is a bijection

$$\begin{aligned} \text{the plane } z = 1 \text{ in } \mathbb{R}^3 &\leftrightarrow \{\text{nonhorizontal lines in } \mathbb{R}^3 \text{ through } O\} \\ P &\mapsto \overleftrightarrow{OP} \\ L \cap \{z = 1\} &\hookleftarrow L. \end{aligned}$$

Think of points in the plane $z = 1$ as *being* the corresponding lines. Extend the plane by introducing new "honorary points" that represent the horizontal lines. They should be thought of as being "points at infinity": for example, as $x \to \infty$, the point $(x, 0, 1)$ tends to infinity in a certain direction, and the corresponding line flattens out and approaches the x-axis.

This yields the projective plane $\mathbb{P}^2(\mathbb{R})$ whose points correspond to arbitrary lines in $\mathbb{R}^3$ through O.

23.2. **Projective space.** Let k be any field. Fix $n \in \mathbb{Z}_{\geq 0}$. Let L be a field extension of k. Define an equivalence relation $\sim$ on $L^{n+1} - \{\vec{0}\}$ such that

$$(a_0, \ldots, a_n) \sim (b_0, \ldots, b_n)$$

if and only if there exists $\lambda \in L^\times$ such that $b_i = \lambda a_i$. Define

$$\mathbb{P}^n_k(L) := \frac{L^{n+1} - \{\vec{0}\}}{\sim}.$$

Here $\mathbb{P}^n_k$ is called n-dimensional projective space over k.

For $a_0, \ldots, a_n \in L$ not all 0, let $(a_0 : \ldots : a_n)$ denote the equivalence class of $(a_0, \ldots, a_n)$. The a_i are called homogeneous coordinates of the point.

The group G_k acts on $\mathbb{P}^n(\overline{k})$, and the fixed subset is $\mathbb{P}^n(k)$. (This will be assigned for homework.)

23.3. **Projective varieties.** It does not make sense to evaluate a polynomial $f \in k[x_0, \ldots, x_n]$ at a point in $\mathbb{P}^n(L)$, because the polynomial has different values at the different representatives of the equivalence class. But if f is homogeneous of some degree d, meaning that in every monomial of f the exponents of the variables sum to d, then the condition that f be 0 at a particular point in $\mathbb{P}^n(L)$ makes sense, since multiplying the homogeneous coordinates by λ multiples the value of f by λ^d.

Given a set T of homogeneous polynomials in $k[x_1, \ldots, x_n]$, define $Z = Z_T$ by the rule

$$Z(L) := \{P \in \mathbb{P}^n(L) : f(P) = 0 \text{ for all } f \in T\}.$$

Any such Z is called an projective variety over k.

An ideal I generated by a set T of homogeneous polynomials is called a homogeneous ideal, and then $Z_I := Z_T$ satisfies

$$Z_I(L) = \{P \in \mathbb{P}^n(L) : f(P) = 0 \text{ for all homogeneous } f \in I\}.$$

Conversely, given a projective variety Z in $\mathbb{P}^n$, its homogeneous ideal I is the ideal generated by the set of homogeneous polynomials f such that $f(P) = 0$ for all $P \in Z(L)$ for all L.

Theorem 23.1. *There is an inclusion-reversing bijection*

$$\{\text{radical homogeneous ideals of } k[x_0, \ldots, x_n] \text{ not } (x_0, \ldots, x_n)\} \leftrightarrow \{\text{projective varieties } Z \text{ in } \mathbb{P}^n_k\}$$
$$I \mapsto Z_I$$
$$\text{homogeneous ideal of } Z \hookleftarrow Z.$$

If $I \leftrightarrow Z$, then

$$S(I) := \frac{k[x_0, \ldots, x_n]}{I}$$

is called the homogeneous coordinate ring and one writes

$$Z = \operatorname{Proj} \frac{k[x_0, \ldots, x_n]}{I}.$$

Definition 23.2. A projective variety is irreducible if it satisfies any of the following equivalent conditions:

- It cannot be written as a union of two smaller projective varieties.
- Its homogeneous ideal is a prime ideal in $k[x_0, \ldots, x_n]$.
- Its homogeneous coordinate ring is an integral domain.

23.4. **Projective varieties as a union of affine varieties.**

23.4.1. *The standard covering of projective space.* There are inclusions

$$\mathbb{A}^2 \hookrightarrow \mathbb{P}^2$$
$$(x, y) \mapsto (x : y : 1)$$

and

$$\mathbb{P}^1 \hookrightarrow \mathbb{P}^2$$
$$(x : y) \mapsto (x : y : 0)$$

These copies of $\mathbb{A}^2$ and $\mathbb{P}^1$ in $\mathbb{P}^2$ are complements of each other. (If a point $(x : y : z) \in \mathbb{P}^2(L)$ has $z \neq 0$, the homogeneous coordinates can be scaled in a unique way to get a point of the form $(x : y : 1)$.)

More generally, inside $\mathbb{P}^n$, if $i \in \{0, 1, \ldots, n\}$, then the hyperplane H_i defined by $x_i = 0$ is a copy of $\mathbb{P}^{n-1}$, and its complement U_i, which consists of points of the form $(x_0 : \cdots : x_{i-1} : 1 : x_{i+1} : \cdots : x_n)$, is a copy of $\mathbb{A}^n$.

Since every point on $\mathbb{P}^n$ has at least one nonzero coordinate, $\bigcup_{i=0}^n U_i = \mathbb{P}^n$.

23.4.2. *Homogenization and dehomogenization of polynomials.* Given a polynomial $f(x, y) \in k[x, y]$, we can make a homogeneous polynomial by multiplying each monomial by a suitable power of z. For example, $5x^2 + 3y^3 + xy + 7$ becomes $5x^2z + 3y^3 + xyz + 7z^3$. The process can be reversed by setting $z = 1$.

In general:

Definition 23.3. Fix $i \in \{0, 1, \ldots, n\}$. Given $f \in k[x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]$ of total degree d, its homogenization is

$$x_i^d f\left(\frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_n}{x_i}\right).$$

Conversely, given a homogeneous polynomial $F(x_0, \ldots, x_n)$, its dehomogenization (with respect to x_i) is

$$F(x_0, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n).$$

23.4.3. *Affine patches of a projective variety.* Let X be a projective variety in $\mathbb{P}^n$. Let $I \subseteq k[x_0, \ldots, x_n]$ be its homogeneous ideal. Fix $i \in \{0, 1, \ldots, n\}$. Let I_i be the ideal of $k[x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]$ obtained by dehomogenizing all homogeneous $f \in I$. Then the i^{th} affine patch of X is the affine variety $X \cap U_i = \operatorname{Spec} \frac{k[x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]}{I_i}$. We have $\bigcup_{i=0}^n (X \cap U_i) = X$.

One thinks of X as being constructed by glueing the affine patches in a particular way. (More general varieties and schemes can be constructed by glueing affine varieties in other ways.)

23.4.4. *Projective closure of an affine variety.* Let $V = \operatorname{Spec} \frac{k[x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]}{I}$ be an affine variety. So $V \subseteq \mathbb{A}^n = U_i \subset \mathbb{P}^n$. The projective closure $\overline{V}$ of V in $\mathbb{P}^n$ is the projective variety defined by the homogeneous ideal generated by the homogenizations of the $f \in I$.

If I is generated by one element, it suffices to homogenize that one element.

Example 23.4. The projective closure of the affine plane curve $y^2 = x^3 + 2x + 7$ in $\mathbb{P}^2$ is the projective variety defined by $y^2z = x^3 + 2xz^2 + 7z^3$.

If one starts with an affine variety V and takes its projective closure, one can recover V by taking an affine patch.

But if one starts with a projective variety X, and takes an affine patch $X \cap U_i$, and then takes the projective closure, one could get a smaller variety: one loses irreducible components in the hyperplane H_i.

23.4.5. *Properties of projective varieties.*

Definition 23.5. The dimension of a projective variety is the maximum of the dimensions of its affine patches.

Definition 23.6. A projective variety is smooth if and only if all its affine patches are.

In fact, one can check whether a point P on a projective variety is singular by checking *any* affine patch containing P.

Definition 23.7. The function field of an irreducible projective variety is the function field of any of its nonempty affine patches. (One can show that this is independent of the patch chosen.)

24. Morphisms and rational maps

Definition 24.1. Let X be an irreducible variety, and let Y be a projective variety in $\mathbb{P}^n$. A rational map $f\colon X \dashrightarrow Y$ is an equivalence class of $(n+1)$-tuples

$$(f_0 : f_1 : \cdots : f_n)$$

such that $f_i \in \kappa(X)$ for all i, and the f_i are not all identically 0, and such that for any field extension $L \supseteq k$ and any $P \in X(L)$ such that the $f_i(P)$ are all defined and not all 0,

$$(f_0(P) : f_1(P) : \cdots : f_n(P)) \in Y(L).$$

The equivalence relation is:

$$(f_0 : f_1 : \cdots : f_n) = (\lambda f_0 : \cdots : \lambda f_n)$$

for any $\lambda \in \kappa(X)^\times$. Say that f is defined (or regular at a point $P \in X(L)$ if there exists $\lambda \in \kappa(X)^\times$ such that

$$(f_0(P) : f_1(P) : \cdots : f_n(P))$$

is defined (i.e., each f_i is defined at P functions the $f_i(P)$ are all defined

Definition 24.2. A rational map $X \dashrightarrow Y$ that is defined at every $P \in X(L)$ (for all $L \supseteq k$) is called a morphism.

Example 24.3. The map

$$\mathbb{P}^1 \to \mathbb{P}^2$$
$$(x : y) \mapsto (x^2 : xy : y^2)$$

is a morphism. (Strictly speaking, it should be written as $(t^2 : t : 1)$ or $(1 : t^{-1} : t^{-2})$, where t is the rational function x/y on $\mathbb{P}^1$.) Its image is the projective curve in $\mathbb{P}^2$ defined by $x_1^2 = x_0 x_2$.

Example 24.4. Consider the unit circle $X\colon x^2+y^2=1$ over a field k of characteristic not 2. Let $\overline{X}$ be its projective closure. Identify $\mathbb{P}^1$ with the projective closure of the y-axis. For all points $P \in X(L)$ other than $(-1,0)$, the line through $(-1,0)$ and P intersects this $\mathbb{P}^1$ in a point $Q \in \mathbb{P}^1(L)$. This construction defines a rational map

$$\begin{aligned} f\colon \overline{X} &\to \mathbb{P}^1 \\ (x:y:1) &\mapsto \left(\frac{y}{x+1}:1\right). \end{aligned}$$

There is an inverse construction: For most points $Q \in \mathbb{P}^1(L)$, the line through $(-1,0)$ and Q intersects X in one point P other than $(-1,0)$, and this defines a rational map

$$\begin{aligned} g\colon \mathbb{P}^1 &\to \overline{X} \\ (t:1) &\to \left(\frac{1-t^2}{1+t^2}:\frac{2t}{1+t^2}:1\right). \end{aligned}$$

Where are these rational maps defined? The first map can be rewritten as

$$(x:y:z) \mapsto (y:x+z) = (x-z:-y).$$

The first right hand side makes sense except at $(1:0:-1)$, and the second right hand side makes sense except at $(1:0:1)$, so it is defined everywhere.

The second map can be rewritten as

$$(x:y) \mapsto (x^2-y^2:2xy:x^2+y^2).$$

which is defined everywhere since $x^2-y^2=x^2+y^2=0$ implies $x=y=0$.

The composition of the two rational maps in either order is the identity map, so one says that the two varieties X and $\mathbb{P}^1$ are isomorphic: $X \simeq \mathbb{P}^1$. In particular, for each field extension $L \supseteq k$, the set $X(L)$ can be parametrized.

Taking $L=\mathbb{Q}$ gives essentially the well-known parametrization of Pythagorean triples.

Remark 24.5. Sometimes it happens that there are rational maps $X \dashrightarrow Y$ and $Y \dashrightarrow X$ whose composition in either order is the identity except that one or both of the maps is not defined everywhere. In this case, X and Y are said to be birational, which is weaker than being isomorphic.

25. Quadratic forms

In this section, k is a field of characteristic not 2.

Definition 25.1. A quadratic form over a field k is a homogeneous polynomial $q(x_1,\dots,x_n) \in k[x_1,\dots,x_n]$ of degree 2.

Example 25.2. Over $\mathbb{Q}$, take $q(x,y)=2x^2+5xy-6y^2$.

A quadratic form gives rise to a function $q\colon V \to k$, where $V = k^n$. Since $\#k > 2$, the function determines the quadratic form, so they will be identified from now on.

More abstractly:

Definition 25.3. A quadratic form on a finite-dimensional k-vector space V is a function $q\colon V \to k$ such that for a choice of basis $e_1, \ldots, e_n$ of V, the function $q(x_1e_1 + \cdots + x_ne_n)$ from $k^n \to k$ is given by a quadratic form in the previous sense.

Definition 25.4. A bilinear form on a k-vector space V is a function

$$B\colon V \times V \to k$$

such that the identities $B(v_1+v_2, w) = B(v_1, w)+B(v_2, w)$, $B(\lambda v, w) = \lambda B(v, w)$, $B(v, w_1+w_2) = B(v, w_1) + B(v, w_2)$, and $B(v, \lambda w) = \lambda B(v, w)$ hold (where $\lambda \in k$ and everything else is a vector in V). A bilinear form is symmetric if

$$B(v, w) = B(w, v)$$

for all $v, w \in V$.

For each V, there is a bijection

$$\begin{aligned} \{\text{quadratic forms on } V\} &\to \{\text{symmetric bilinear forms on } V\} \\ q &\mapsto B(x, y) := \frac{q(x+y) - q(x) + q(y)}{2} \\ q(x) := B(x, x) &\mapsfrom B. \end{aligned}$$

These can also be described in matrix form: $q(x) = x^tAx$ and $B(x, y) = x^tAy$ for a unique symmetric matrix A; here x and y are viewed as column vectors, and x^t denotes the transpose (a row vector).

Definition 25.5. The rank of a quadratic form is the rank of the associated symmetric matrix A.

Definition 25.6. The quadratic form $q(x_1, \ldots, x_n)$ is called nondegenerate if any of the following equivalent conditions hold:

- The associated symmetric matrix A is invertible.
- For each nonzero $x \in V$, the linear map $y \mapsto B(x, y)$ is nonzero.
- The rank of q equals n.

25.1. **Equivalence of quadratic forms.**

Definition 25.7. Two quadratic forms $q(x_1, \ldots, x_n)$ and $q'(x_1, \ldots, x_n)$ are equivalent if they differ by an linear change of variable: $q'(x) = q(Tx)$ for some invertible matrix T.

Example 25.8. Are x^2+y^2 and $5x^2+5y^2$ equivalent over $\mathbb{Q}$? Answer: Yes, because

$$(2x+y)^2+(x-2y)^2=5x^2+5y^2.$$

How about x^2+y^2 and $3y^2+3z^2$? It turns out that this time the answer is no.

It is not so easy to tell when two quadratic forms are equivalent!

Proposition 25.9. *Every quadratic form $q(x_1,\dots,x_n)$ over k is equivalent to a diagonal quadratic form*

$$a_1x_1^2+a_2x_2^2+\cdots+a_nx_n^2.$$

Proof. We use induction on $\dim V$. The cases $\dim V\le 1$ are trivial.

If $q\equiv 0$, then just take all a_i to be 0. Otherwise choose v with $q(v)\neq 0$. Since $x\mapsto B(x,v)$ is a surjective linear map $V\to k$, its kernel $v^\perp:=\{x\in V: B(x,v)=0\}$ is of dimension 1 less than V. Also, $v\notin v^\perp$ (since $q(v)=B(v,v)\neq 0$), so $V\simeq kv\oplus v^\perp$. If $y=y_1+y_2$ with $y_1\in kv$ and $y_2\in v^\perp$, then $q(y)=q(y_1)+q(y_2)+2B(y_1,y_2)=q(y_1)+q(y_2)$. By the inductive hypothesis, $q|_{v^\perp}$ can be diagonalized, and $q(x_1v)$ is of the form $a_1x_1^2$, where $a_1=q(v)$. □

Remark 25.10. If q is equivalent to $a_1x_1^2+\cdots+a_nx_n^2$, then the rank of q equals the number of nonzero a_i.

25.2. **Numbers represented by quadratic forms.**

Definition 25.11. Let q be a quadratic form on V, and let $a\in k$. Say that q represents a if there exists a nonzero $x\in V$ such that $q(x)=a$.

The condition that x be nonzero matters only when $a=0$. In this case it is important to include this in the definition, since otherwise every quadratic form would represent 0!

Example 25.12. The quadratic form x^2-2y^2 over $\mathbb{Q}$ represents -7 but not 0.

Proposition 25.13. *If a nondegenerate quadratic form q represents 0, then it represents every element of k.*

Proof. Choose $e\in V$ such that $q(e)=0$. Since q is nondegenerate, there exists $f\in V$ with $B(e,f)\neq 0$, and f must be independent of e. Then $q(xe+yf)=axy+by^2=(ax+by)y$ for some $a,b\in k$ with $a=2B(e,f)\neq 0$. For any $c\in k$, we can solve $(ax+by)y=c$ by setting $y=1$ and solving a linear equation for x. Thus even q restricted to the subspace $\langle e,f\rangle$ represents c. □

26. Local-global principle for quadratic forms

Theorem 26.1 (Hasse-Minkowski)**.** *A quadratic form over $\mathbb{Q}$ represents 0 if and only if it represents 0 over $\mathbb{Q}_p$ for all $p\le\infty$.*

(Actually, this was proved by Minkowski alone. Hasse generalized the theorem to the case of quadratic forms over a finite extension of $\mathbb{Q}$.)

Remark 26.2. The fields $\mathbb{Q}_p$ and $\mathbb{R}$ and $\mathbb{F}_p((t))$ and their finite extensions are called local fields, because Laurent series fields like $\mathbb{C}((t))$ are describing the expansion of functions around a single point. On the other hand, $\mathbb{Q}$ and $\mathbb{F}_p(t)$ and their finite extensions are called global fields. Local fields are completions of global fields.

Here are two variants of the theorem:

Theorem 26.3. *Given $a \in \mathbb{Q}$, a quadratic form over $\mathbb{Q}$ represents a if and only if it represents a over $\mathbb{Q}_p$ for all $p \leq \infty$.*

Theorem 26.4. *Two quadratic forms over $\mathbb{Q}$ are equivalent if and only if they are equivalent over $\mathbb{Q}_p$ for all $p \leq \infty$.*

Corollary 26.5. *Let X be a (smooth projective) plane conic over $\mathbb{Q}$ (i.e., the zero locus in $\mathbb{P}^2_{\mathbb{Q}}$ of a quadratic form $q(x, y, z)$ that is irreducible even over $\overline{\mathbb{Q}}$). Then the following are equivalent:*

(i) *X has a rational point.*
(ii) *X has a $\mathbb{Q}_p$-point for all $p \leq \infty$.*
(iii) *$X \simeq \mathbb{P}^1_{\mathbb{Q}}$.*

Proof. (i) $\iff$ (ii) is Hasse-Minkowski.
(iii) $\implies$ (i) is trivial.
(i) $\implies$ (iii): If X has a rational point P, projection from P defines an isomorphism (the argument is similar to the argument for the unit circle). $\square$

Remark 26.6. If X is a smooth projective plane conic over $\mathbb{F}_q$ then X has an $\mathbb{F}_q$-point, by the Chevalley-Warning theorem proved in the homework, so $X \simeq \mathbb{P}^1_{\mathbb{F}_q}$. In particular $\#X(\mathbb{F}_q) = q + 1$.

Definition 26.7. A variety X over $\mathbb{Q}$ is said to satisfy the local-global principle (also called the Hasse principle) if the implication

$$X \text{ has a } \mathbb{Q}_p\text{-point for all } p \leq \infty \implies X \text{ has a } \mathbb{Q}\text{-point}$$

holds.

So plane conics satisfy the local-global principle. Unfortunately, more complicated varieties can violate the local-global principle. It is a major problem of arithmetic geometry to determine which families of varieties satisfy the local-global principle.

26.1. **Proof of the Hasse-Minkowski theorem for quadratic forms in 2 or 3 variables.** Note: In prove the Hasse-Minkowski theorem, we can assume that the quadratic form is in diagonal form, and that the first coefficient is 1 (scaling it by a nonzero constant does not affect whether it represents 0). We do only the hard direction, in which we assume that q represents 0 over $\mathbb{Q}_p$ for all $p \leq \infty$, and hope to prove that q represents 0 over $\mathbb{Q}$.

First consider the 2-variable case, so q is $x^2 - ay^2$ for some $a \in \mathbb{Q}$. To say that $x^2 - ay^2$ represents 0 is to say that a is a square. We may assume $a \neq 0$. Since q represents 0 over $\mathbb{R}$, we have $a > 0$. Write

$$a = \prod_{\text{primes p}} p^{n_p}.$$

Since q represents 0 over $\mathbb{Q}_p$, the valuation n_p must be even. Since this holds for all p, this means that a is a square in $\mathbb{Q}$, so q represents 0.

The proof in the 3-variable case will use the following lemma.

Lemma 26.8. *Let $a, b \in k$ where $\operatorname{char} k \neq 2$. Let $N\colon k(\sqrt{a}) \to k$ be the norm map: if a is not a square in k, then $N(x + y\sqrt{a}) = x^2 - ay^2$. (If a is a square, $N(x) := x$.) Then the quadratic form $x^2 - ay^2 - bz^2$ over k represents 0 if and only if $b = N(\alpha)$ for some $\alpha \in k(\sqrt{a})$.*

Proof. Case 1: a is a square, say $a = c^2$. Then $x^2 - ay^2 = (x + cy)(x - cy)$, which is equivalent to xy, which represents everything, so $x^2 - ay^2 - bz^2 = 0$ has a solution with $z = 1$. On the other side, $b = N(b)$.

Case 2: a is not a square. If b is a norm, say $b = N(x + y\sqrt{a})$, then $x^2 - ay^2 - b \cdot 1^2 = 0$. Conversely, if $x^2 - ay^2 - bz^2$ represents 0, the nontrivial solution to $x^2 - ay^2 - bz^2 = 0$ must have $z \neq 0$. Dividing by z^2 shows that b is a norm. □

We may assume that our 3-variable quadratic form q is $x^2 - ay^2 - bz^2$ where $a, b \neq 0$. Multiplying y or z by an element of $\mathbb{Q}^\times$ changes q to an equivalent quadratic form, so we are free to multiply a and b by squares. Thus we may assume that a and b are integers, and in fact, squarefree integers (i.e., not divisible by the square of any prime).

We use strong induction on $m := |a| + |b|$.

Case 1: $m \leq 2$. There are four possibilities:

$$x^2 + y^2 + z^2$$
$$x^2 + y^2 - z^2$$
$$x^2 - y^2 + z^2$$
$$x^2 - y^2 - z^2.$$

We are assuming that q represents 0 over $\mathbb{R}$, so the first is actually not possible. In the other three cases, q represents 0 over $\mathbb{Q}$, as desired.

Case 2: $m > 2$. Without loss of generality $|b| \geq |a|$. So $|b| \geq 2$. Write

$$b = \pm p_1 \cdots p_k$$

where the p_i are distinct primes. Let p be one of the p_i. By assumption, there is a nontrivial solution to $x^2 - ay^2 - bz^2 = 0$ over $\mathbb{Q}_p$, and we may assume that $x, y, z \in \mathbb{Z}_p$ and that not all are in $p\mathbb{Z}_p$.

We claim that a is a square mod p. If not, then considering $x^2 - ay^2 - bz^2 = 0$ modulo p shows that $x \equiv y \equiv 0 \pmod{p}$, but then p^2 divides x^2 and ay^2, so $p^2|bz^2$, so $p|z^2$, so $p|z$, so $x, y, z \in p\mathbb{Z}_p$, a contradiction.

Since a is a square mod p_i for all i, and since $\mathbb{Z}/b\mathbb{Z} = \prod \mathbb{Z}/p_i\mathbb{Z}$, we have that a is a square mod b. So there exists $t \in \mathbb{Z}$ such that $t^2 \equiv a \pmod{b}$. Adjust t by a multiple of b to assume that $|t| \leq |b|/2$. So

$$t^2 - a = bb'$$

for some $b' \in \mathbb{Z}$. We have

$$|b'| = \left|\frac{t^2 - a}{b}\right| \leq \frac{|t|^2}{|b|} + \frac{|a|}{|b|} \leq \frac{|b|}{4} + 1 < |b|$$

since $|b| \geq 2$.

Now bb' is a norm of an element of $\mathbb{Q}(\sqrt{a})$, and hence is a norm from $\mathbb{Q}_p(\sqrt{a})$. Lemma 26.8 implies that b too is a norm from $\mathbb{Q}_p(\sqrt{a})$, so $b' = (bb')/b$ is a norm from $\mathbb{Q}_p(\sqrt{a})$. Thus

$$x^2 - ay^2 - b'z^2 = 0$$

represents 0 over each $\mathbb{Q}_p$. But $|a| + |b'| < |a| + |b|$ (and it's even better if you divide b' by a square to get a squarefree coefficient), so the inductive hypothesis implies that it represents 0 over $\mathbb{Q}$. Thus b' is a norm from $\mathbb{Q}(\sqrt{a})$. If $b' = 0$, then a is a square, and we are done; otherwise $b = (bb')/b'$ is a norm from $\mathbb{Q}(\sqrt{a})$, and Lemma 26.8 implies that $x^2 - ay^2 - bz^2 = 0$ represents 0.

27. Rational points on conics

Consider a projective plane conic $ax^2 + by^2 + cz^2 = 0$ in $\mathbb{P}^2_{\mathbb{Q}}$. Without loss of generality, a, b, c are nonzero integers.

Proposition 27.1. *If $a, b, c \in \mathbb{Z}$ are all nonzero, and p is a finite prime such that $p \nmid 2abc$, then $ax^2 + by^2 + cz^2 = 0$ has a nontrivial solution over $\mathbb{Q}_p$.*

Proof. By the Chevalley-Warning theorem, there exists a nontrivial solution over $\mathbb{F}_p$. Lift this solution arbitrarily to get $(x_0, y_0, z_0) \in \mathbb{Z}_p$ satisfying $ax_0^2 + by_0^2 + cz_0^2 \equiv 0 \pmod{p}$, with x_0, y_0, z_0 not all in $p\mathbb{Z}_p$. Without loss of generality, suppose that $x_0 \notin p\mathbb{Z}_p$. Then x_0 is an approximate zero of the polynomial

$$f(x) := ax^2 + by_0^2 + cz_0^2 \in \mathbb{Z}_p[x]$$

and $p \nmid f'(x_0) = 2ax_0$, so Hensel's lemma gives an exact solution $x_1 \in \mathbb{Z}_p$ to $f(x_1) = 0$ with $x_1 \equiv x_0 \not\equiv 0 \pmod{p\mathbb{Z}_p}$. So (x_1, y_0, z_0) is a solution to $ax^2 + by^2 + cz^2 = 0$. □

Example 27.2. For which $p \leq \infty$ does $x^2 + y^2 = 3z^2$ in $\mathbb{P}^2_{\mathbb{Q}}$ have a $\mathbb{Q}_p$-point?

According to Proposition 27.1, it automatically has a $\mathbb{Q}_p$-point for all p except possibly $\infty, 2, 3$. It has the $\mathbb{R}$-point $(\sqrt{3} : 0 : 1)$.

If there were a $\mathbb{Q}_3$-point, it would have the form $(x : y : z)$ with $x, y, z \in \mathbb{Z}_3$ not all divisible by 3. Considering the equation modulo 3 yields $x^2 + y^2 \equiv 0 \pmod 3$, which implies $x \equiv y \equiv 0 \pmod 3$ since -1 is not a square in $\mathbb{F}_3$. But then 3^2 divides $x^2 + y^2 = 3z^2$, so $3|z$, contradicting the assumption that not all of x, y, z are divisible by 3. Thus there is no $\mathbb{Q}_3$-point.

Similarly, a $\mathbb{Q}_2$-point would have the form $(x : y : z)$ with $x, y, z \in \mathbb{Z}_2$ not all divisible by 2. Then $0 = x^2 + y^2 - 3z^2 \equiv x^2 + y^2 + z^2 \pmod 4$, but squares in $\mathbb{Z}_2$ are 0 or 1 mod 4, so $x^2 + y^2 + z^2$ can be 0 mod 4 only if $2|x, y, z$, a contradiction. Thus there is no $\mathbb{Q}_2$-point.

Remark 27.3. Using quadratic reciprocity, one can show that for any smooth conic X over $\mathbb{Q}$, the number of $p \leq \infty$ such that X has no $\mathbb{Q}_p$-point is finite and *even*!

28. Sums of three squares

Lemma 28.1. *A nonzero rational number a is represented by $x^2 + y^2 + z^2$ over $\mathbb{Q}$ if and only if $a > 0$ and a is not of the form $4^m u$ with $u \in 7 + 8\mathbb{Z}_2$.*

Proof. Since $x^2 + y^2 + z^2$ represents 0 over $\mathbb{Q}_p$ for all odd primes p, it also represents a over such $\mathbb{Q}_p$. It also represents a over $\mathbb{R}$ since $a > 0$. So the question is whether it represents a over $\mathbb{Q}_2$.

First consider the range of $x^2 + y^2 + z^2$ where $x \in \mathbb{Z}_2^\times$ and $y, z \in \mathbb{Z}_2$. The range of x^2 is $1 + 8\mathbb{Z}_2$, so the range of $x^2 + y^2 + z^2$ is a union of cosets of $8\mathbb{Z}_2$, and we just try all possibilities modulo 8. Namely, y^2 is 0, 1, or 4 modulo 8, and z^2 is similar, so the range of $x^2 + y^2 + z^2$ is $\{1, 2, 3, 5, 6\} + 8\mathbb{Z}_2$. A general triple $(x, y, z) \in (\mathbb{Q}_2)^3 - \{\vec{0}\}$ is obtained from one as above by multiplying by a power of 2 (and permuting the variables), and this multiplies the output by a power of 4. □

Theorem 28.2 (Gauss)**.** *A positive integer a is a sum of three integer squares if and only if it is not of the form $4^m(8n + 7)$ with $m, n \in \mathbb{Z}_{\geq 0}$.*

Idea of proof, following Davenport and Cassels. By Lemma 28.1, it suffices to show that for $a \in \mathbb{Z}_{>0}$, if $x^2 + y^2 + z^2 = a$ has a rational solution, then it has an integer solution. The idea is this: Given a rational point P on the sphere $x^2 + y^2 + z^2 = a$, let Q be the nearest point with integer coordinates. If $P = Q$, we are done. Otherwise $\overrightarrow{PQ}$ intersects the sphere in another rational point R, and the fact that $PQ \leq \sqrt{(1/2)^2 + (1/2)^2 + (1/2)^2} < 1$ implies

(with some work) that $\mathrm{denom}(R) < \mathrm{denom}(P)$, where $\mathrm{denom}(P)$ denotes the lcm of the denominators of the coordinates of P. □

The three squares theorem has two nice corollaries.

Corollary 28.3 (Lagrange)**.** *Every $a \in \mathbb{Z}_{\geq 0}$ is a sum of four squares.*

Proof. If a is a sum of three squares, let the fourth square be 0^2. Otherwise $a = 4^m(8n+7)$ for some $m, n \in \mathbb{Z}_{\geq 0}$. Write $8n+6$ as a sum of three squares; then $8n+7$ is a sum of four squares, and the same is true of a. □

Corollary 28.4 (Gauss)**.** *Every $a \in \mathbb{Z}_{\geq 0}$ is a sum of three triangular numbers (i.e., three numbers of the form $m(m+1)/2$).*

Proof. The key trick: if $x = 2m+1$, then

$$\frac{x^2-1}{8} = \frac{m(m+1)}{2}.$$

By the three squares theorem, $8a+3$ is a sum of three squares:

$$x_1^2 + x_2^2 + x_3^2 = 8a+3.$$

Considering this equation modulo 4 shows that x_1, x_2, x_3 are all odd. Write $x_i = 2m_i + 1$. Then

$$\frac{m_1(m_1+1)}{2} + \frac{m_2(m_2+1)}{2} + \frac{m_3(m_3+1)}{2} = a.$$

□

29. Valuations on the function field of a curve

Definition 29.1. A curve is a 1-dimensional variety.

Let C be a curve over k. Let $\kappa(C)$ be the function field of C.

Definition 29.2. Let $P \in C(k)$. Suppose that C is smooth at P. The local ring $\mathcal{O}_P$ of C at P is the set of functions $f \in \kappa(C)$ that are regular (defined) at P. Let $\mathfrak{m}_P := \{f \in \mathcal{O}_P : f(P) = 0\}$, which is a maximal ideal of $\mathcal{O}_P$.

Example 29.3. Take $C = \mathbb{A}^1_k$. Let P be the origin. Then

$$\kappa(C) = \left\{ \frac{p(t)}{q(t)} : p(t), q(t) \in k[t] \text{ and } q(t) \text{ is not the zero polynomial} \right\} =: k(t)$$

$$\mathcal{O}_P = \left\{ \frac{p(t)}{q(t)} : q(0) \neq 0 \right\}$$

$$\mathfrak{m}_P = \left\{ \frac{p(t)}{q(t)} : p(0) = 0 \text{ and } q(0) \neq 0 \right\}$$

$$\mathcal{O}_P^\times = \left\{ \frac{p(t)}{q(t)} : p(0) \neq 0 \text{ and } q(0) \neq 0 \right\}.$$

Every $f \in \kappa(C)^\times$ can be uniquely written as $t^n u$ where $n \in \mathbb{Z}$ and $u \in \mathcal{O}_P^\times$. The map

$$\begin{aligned} v_P \colon \kappa(C) &\to \mathbb{Z} \cup \{+\infty\} \\ f = t^n u &\mapsto n \\ 0 &\mapsto +\infty \end{aligned}$$

is a valuation on $\kappa(C)$. We have

$$\begin{aligned} \mathcal{O}_P &= \{f \in \kappa(C) : v_P(f) \geq 0\} \\ \mathfrak{m}_P &= \{f \in \kappa(C) : v_P(f) > 0\} \\ \mathcal{O}_P^\times &= \{f \in \kappa(C) : v_p(f) = 0\}. \end{aligned}$$

Also, $v_P(t) = 1$.

All of this generalizes to any smooth curve.

Theorem 29.4. *Let C be a smooth curve. Let $P \in C(k)$. Then there is a valuation*

$$v_P \colon \kappa(C) \to \mathbb{Z} \cup \{+\infty\}$$

such that

$$\begin{aligned} \mathcal{O}_P &= \{f \in \kappa(C) : v_P(f) \geq 0\} \\ \mathfrak{m}_P &= \{f \in \kappa(C) : v_P(f) > 0\} \\ \mathcal{O}_P^\times &= \{f \in \kappa(C) : v_p(f) = 0\}. \end{aligned}$$

Definition 29.5. Say that f has a zero of multiplicity m at P if $v_P(f) = m > 0$, and a pole of multiplicity m at P if $v_P(f) = -m < 0$.

Definition 29.6. An element $t \in \kappa(C)$ such that $v_P(t) = 1$ is called a uniformizing parameter at P.

If t is a uniformizing parameter at P, then every $f \in \kappa(C)^\times$ can be uniquely written as $t^n u$, where $n \in \mathbb{Z}$ and $u \in \mathcal{O}_P^\times$. Namely, $n = v_P(f)$.

Over a field like $\mathbb{R}$, the implicit function theorem shows that the part of the curve near P is the graph of an analytic function of t, so the different values of t near $t = 0$ parametrize the points of C near P.

Remark 29.7. Suppose that C is the curve $f(x, y) = 0$ in $\mathbb{A}^2$, and $(a, b) \in C(k)$ is a smooth point on C, so either $\frac{\partial f}{\partial y}(a, b) \neq 0$ or $\frac{\partial f}{\partial x}(a, b) \neq 0$ (or both).

- If $\frac{\partial f}{\partial y}(a, b) \neq 0$ (so the tangent line is not vertical), then $x - a$ is a uniformizing parameter.
- If $\frac{\partial f}{\partial x}(a, b) \neq 0$, then $y - b$ is a uniformizing parameter.

Example 29.8. Let C be the curve $y^2 = x^3 - x$. Let $P = (0, 0)$. At P, the rational function y is a uniformizing parameter. So $v_P(y) = 1$. What is $v_P(x)$? We have $x = y^2 \left(\frac{1}{x^2-1}\right)$, and $\frac{1}{x^2-1} \in \mathcal{O}_P^\times$ (it and its inverse are both defined at P), so $v_P(x) = 2$.

29.1. **Closed points.** If k is not algebraically closed (but still perfect), then we will want to define valuations at more than just the k-points. Going back to the example of $\mathbb{A}^1_k$, the valuation at a k-point was measuring the exponent of $t - a$ in the factorization of a rational function. But we should also measure the exponent of each other monic irreducible polynomial $p(t)$ in $k[t]$. The zero set of any such $p(t)$ is an irreducible subvariety of $\mathbb{A}^1_k$, but when considered over $\overline{k}$ it breaks up as a G_k-orbit of points in $\mathbb{A}^1(\overline{k})$.

In general, a closed point of a variety X is a 0-dimensional irreducible subvariety. If $X = \operatorname{Spec} k[t_1, \ldots, t_n]/I$, then closed points of X are in bijection with maximal ideals of $k[t_1, \ldots, t_n]/I$. If k is algebraically closed, then the closed points are the same as elements of $X(k)$. For an arbitrary perfect field k, the closed points of X are in bijection with the G_k-orbits of points in $X(\overline{k})$.

If P is a closed point of a curve C over k, one can define $\mathcal{O}_P$ and $\mathfrak{m}_P$ as before. The residue field $\kappa(P) := \mathcal{O}_P/\mathfrak{m}_P$ turns out to be a finite extension of k, and $\deg P := [\kappa(P) : k]$ is called the degree of P. If moreover X is a curve C, and C is smooth at P (which is the same as saying that C over $\overline{k}$ is smooth at any of the $\overline{k}$-points into which P breaks up), then there is also a valuation v_P with the same properties as in the case where $P \in C(k)$.

Working with closed points is an alternative to working with L-points for all (finite) extensions L of k.

30. Review

- Absolute values, archimedean vs. nonarchimedean
- Valuations
- Ostrowski's theorem
- Cauchy sequences
- Completion
- $\mathbb{Z}_p$ as inverse limit
- $\mathbb{Q}_p = \operatorname{Frac} \mathbb{Z}_p$, or $\mathbb{Q}_p$ as completion of $\mathbb{Q}$
- Hensel's lemma
- Structure of $\mathbb{Z}_p^\times$ and $\mathbb{Q}_p^\times$
- Squares in $\mathbb{Q}_p^\times$
- p-adic power series
- Algebraic closure
- Finite fields, Frobenius automorphism
- Inverse limits
- Profinite groups, open and closed subgroups

- Properties of fields and extensions of fields: normal, separable, perfect, Galois
- Infinite Galois groups as profinite groups, absolute Galois group
- Infinite Galois theory
- Affine varieties, affine coordinate ring
- Projective varieties, homogeneous coordinate ring
- Irreducibility and function field
- Dimension
- Smoothness
- Homogenization, dehomogenization, projective closure, affine patches
- Rational maps, morphisms
- Quadratic forms, bilinear forms
- Rank, nondegenerate, equivalence
- Local-global principle for quadratic forms (Hasse-Minkowski theorem); applications to rational points on conics
- Valuations on a curve, local ring, maximal ideal, uniformizing parameter

31. Curves and function fields

Theorem 31.1. *If $\phi\colon C \dashrightarrow X$ is a rational map from a smooth irreducible curve to a projective variety, then ϕ is a morphism (i.e., ϕ is actually defined everywhere).*

Proof. It suffices to check that ϕ is defined at each closed point P. Suppose that $X \subseteq \mathbb{P}^n$ and that ϕ is given by $(f_0 : \cdots : f_n)$. Let f be the f_i such that $v_P(f_i)$ is minimum. Then $(f_0 : \cdots : f_n)$ is equivalent to $\left(\frac{f_0}{f} : \cdots : \frac{f_n}{f}\right)$ but $v_p(f_j/f) \geq 0$ for all j so the functions f_j/f are defined at P, and their values are not all 0 since $f_i/f = 1$. So we get a morphism $\phi\colon C \to \mathbb{P}^n$, and in fact it maps into X, because the locus in C where the image satisfies the equations of X in $\mathbb{P}^n$ is a subvariety of C containing infinitely many $\overline{k}$-points. (Every subvariety of C other than C itself is 0-dimensional, and hence a finite union of closed points, which contains only finitely many $\overline{k}$-points.) □

Example 31.2. If C is not smooth, Theorem 31.1 can fail:

$$\begin{aligned} \{y^2 = x^2(x+1)\} &\to \mathbb{A}^1 \\ (x, y) &\mapsto \frac{y}{x} \end{aligned}$$

gives a rational map between the projective closures that is not defined at the singularity $(0, 0)$. Over $\mathbb{R}$, this map cannot even be extended to a *continuous* function.

Example 31.3. Similarly,

$$\{y^2 = x^3\} \to \mathbb{A}^1$$
$$(x, y) \mapsto \frac{y}{x}$$

gives a rational map between the projective closures that is not defined at the singularity $(0, 0)$.

An irreducible curve that is not necessarily smooth or projective is birational to a curve that is smooth and projective. (The analogue for higher-dimensional varieties is an unsolved problem in the case where $\operatorname{char} k > 0$! This is called resolution of singularities.)

Definition 31.4. A variety over k is nice if it is smooth, projective, and geometrically irreducible (i.e., irreducible even when considered over $\overline{k}$). This is not universally accepted terminology, but it is convenient!

Fact: A nonconstant morphism of curves $\phi\colon C' \to C$ defines a field homomorphism

$$\kappa(C) \to \kappa(C')$$
$$f \mapsto f \circ \phi$$

in the opposite direction, and this makes $\kappa(C')$ a *finite* extension of $\kappa(C)$. Define the degree of ϕ to be the degree of this extension, i.e., $\deg \phi := [\kappa(C') : \kappa(C)]$.

Remark 31.5. If k is algebraically closed, then any nonconstant morphism $C' \to C$ induces a surjection $C'(k) \to C(k)$, and for all but finitely many $P \in C(k)$ the number of preimages of P in $C'(k)$ equals the separable degree of $\kappa(C')$ over $\kappa(C)$. In particular, this gives an alternative definition of $\deg \phi$, at least if $\kappa(C')$ is known to be separable over $\kappa(C)$ (for example, if $\operatorname{char} k = 0$).

Say that a field extension K of k is a 1-dimensional function field over k if K is a finite extension of a rational function field $k(t)$ and K contains no nontrivial finite extension of k.

Theorem 31.6. *Then there is an equivalence of categories*

$$\left\{\begin{array}{c}\text{nice curves over } k,\\ \text{nonconstant morphisms}\end{array}\right\} \quad \leftrightarrow \quad \left\{\begin{array}{c}\text{1-dimensional function fields over } k,\\ \text{field homomorphisms acting as the identity on } k\end{array}\right\}^{\text{op}}$$
$$C \quad \mapsto \quad \kappa(C).$$

(The $^{\text{op}}$ *indicates that morphisms give rise to field homomorphisms in the opposite direction.)*

The proof is involved, so we will skip it.

Example 31.7. Let C_0 be the affine curve $y^2 = x^3 - x$ over $\mathbb{Q}$. Its projective closure is a nice curve C. The map

$$C_0 \to \mathbb{A}^1$$
$$(x, y) \mapsto x$$

extends to a morphism

$$C \to \mathbb{P}^1$$

and the reverse map of function fields is the inclusion

$$k(x) \hookrightarrow \operatorname{Frac} \frac{k[x,y]}{(y^2 - x^3 + x)} = k(x,y) = k(x)(\sqrt{x^3 - x}),$$

which is a degree 2 extension of fields.

32. Divisors

Definition 32.1. A divisor is a formal sum $\sum_{\text{closed points } P \in C} n_P P$ such that $n_P \in \mathbb{Z}$ for all P, and $n_P = 0$ for almost all P.

In other words, a divisor on C is a formal integer linear combination of (finitely many) closed points.

Example 32.2. Let C be the projective curve $x^2 + y^2 = z^2$ over $\mathbb{Q}$. Let $P = (1 : 0 : 1)$. Let $Q = (3 : 4 : 5)$. Then $2P - 3Q$ is a divisor.

Definition 32.3. The divisor group $\operatorname{Div} C$ is the group of all divisors on C under addition.

In other words, $\operatorname{Div} C$ is the free abelian group having as basis the set of closed points of C.

There is a partial order on $\operatorname{Div} C$: namely, $\sum n_P P \geq \sum m_P P$ means that $n_P \geq m_P$ for all P.

Definition 32.4. A divisor $D = \sum n_P P$ is called effective if $D \geq 0$ (i.e., $n_P \geq 0$ for all P).

A divisor D can be written as $D_1 - D_2$ where D_1 and D_2 are effective divisors. Moreover, this representation is unique if we also insist that D_1 and D_2 have "disjoint supports".

32.1. **Degree of a divisor.** Recall that if P is a closed point on C, then $\deg P$ is defined as the degree of $\kappa(P) := \mathcal{O}_P/\mathfrak{m}_P$ as a field extension of k.

Definition 32.5. If $D = \sum n_P P$ is a divisor on C, then the degree of D is defined as $\deg D := \sum n_P (\deg P)$.

Example 32.6. Suppose that k is algebraically closed. Then each finite extension $\kappa(P)$ of k must equal k, so each closed point P has degree 1 (they are just the elements of $C(k)$). Thus if $D = \sum n_P P$, then $\deg D = \sum n_P$.

Example 32.7. In the $x^2+y^2=z^2$ example above, the degree of the divisor $2P-3Q$ is -1.

The map

$$\begin{aligned} \operatorname{Div} C &\to \mathbb{Z} \\ D &\mapsto \deg D \end{aligned}$$

is a group homomorphism.

Its kernel, the subgroup of divisors of degree 0, is denoted $\operatorname{Div}^0 C$.

32.2. **Base extension.**

Definition 32.8. If X is a variety over a field k, and L is a field extension of k, then the base extension X_L is the variety defined by the same polynomial equations as X but with the polynomials viewed as polynomials with coefficients in L (even though the coefficients are actually in the subfield k).

A common case is where k is a perfect field and $L=\overline{k}$ is an algebraic closure of k.

Example 32.9. If $X = \operatorname{Spec} \frac{\mathbb{Q}[x,y]}{(x^2+y^2-1)}$, then $X_{\overline{\mathbb{Q}}} = \operatorname{Spec} \frac{\overline{\mathbb{Q}}[x,y]}{(x^2+y^2-1)}$. Similarly, if $Y = \operatorname{Proj} \frac{\mathbb{Q}[x,y,z]}{(x^2+y^2-z^2)}$, then $Y_{\overline{\mathbb{Q}}} = \operatorname{Proj} \frac{\overline{\mathbb{Q}}[x,y,z]}{(x^2+y^2-z^2)}$.

If P is a closed point of C, then its base extension (to $\overline{k}$) consists of a finite set of closed points $P_1,\dots,P_n$ of $C_{\overline{k}}$, where $n=\deg P$. Define a homomorphism

$$\operatorname{Div} C \to \operatorname{Div} C_{\overline{k}}$$

by mapping each closed point P of C to the corresponding sum $P_1+\ldots+P_n$, and extending linearly (i.e., extend so as to get a homomorphism).

Example 32.10. Suppose that C is $\mathbb{P}^1_{\mathbb{R}}$. Then $C_{\mathbb{C}}$ is $\mathbb{P}^1_{\mathbb{C}}$. Closed points on C other than the point $(1:0)$ "at infinity" are closed points in $\mathbb{A}^1_{\mathbb{R}}$, which correspond to monic irreducible polynomials in $\mathbb{R}[t]$. Each such polynomial has degree 1 or 2, and that degree is the degree of the closed point. The base extension of a closed point other than $(1:0)$ is a set of 1 or 2 points in $C(\mathbb{C})$ corresponding to the zeros of the monic irreducible polynomial.

Proposition 32.11. *The homomorphism*

$$\operatorname{Div} C \to \operatorname{Div} C_{\overline{k}}$$

is injective, and its image is the subgroup of G_k-invariant elements of $\operatorname{Div} C_{\overline{k}}$.

Sketch of proof. This follows from the description of a closed point of C as a G_k-orbit of elements of $C(\overline{k})$. □

Example 32.12. Let C be $x^2+y^2=1$ over $\mathbb{Q}$. Let $P=(1/2,\sqrt{3}/2)$ and $Q=(1/2,\sqrt{3}/2)$; these are points in $C(\overline{\mathbb{Q}})$. Even though P and Q are not individually elements of $C(\mathbb{Q})$, their sum $P+Q$ is a $G_{\mathbb{Q}}$-invariant divisor, so it comes from a closed point of C. Namely, it comes from the closed point defined by the equations $x^2+y^2=1$ and $x=1/2$, that is, the closed point $\operatorname{Spec}\frac{\mathbb{Q}[x.y]}{(x^2+y^2-1,x-1/2)}$. This is a closed point of degree 2, with residue field $\mathbb{Q}(\sqrt{3})$.

32.3. **Principal divisors.** Suppose that C is a nice curve over k. Let $f\in\kappa(C)^\times$ be a rational function on C. Then the divisor of f is the divisor

$$\operatorname{div} f=(f):=\sum_{\text{closed points } P\in C} v_P(f)P.$$

Implicit in this definition is the proposition (which we assume without proof) that for any $f\in\kappa(C)^\times$, there are only finitely many P such that $v_P(f)=0$.

Definition 32.13. A divisor is called principal if it equals (f) for some $f\in\kappa(C)^\times$.

The map

$$\begin{aligned}\kappa(C)^\times &\to \operatorname{Div} C\\ f &\mapsto (f)\end{aligned}$$

is a homomorphism, and its image is the set of principal divisors. This shows that the set of principal divisors is a subgroup of $\operatorname{Div} C$.

Example 32.14. If $C=\mathbb{P}^1_k$, then $\kappa(C)$ is the rational function field $k(t)$. Let P be a closed point of C, and let $p(t)$ be the corresponding monic irreducible polynomial. If $f\in\kappa(C)^\times$, then $v_P(f)$ is measuring the exponent of $p(t)$ in f. Thus the divisor of f is keeping track of the complete factorization of f. In other words it measures the zeros and poles of f with multiplicity, with poles giving a negative coefficient.

Remark 32.15. For any rational function $f\in\kappa(C)^\times$, if we write the principal divisor (f) as D_1-D_2 where D_1 and D_2 are effective with disjoint supports, then the following positive integers are equal:

- The degree of the rational map $C\to\mathbb{P}^1$ given by $(f:1)$;
- $\deg D_1$, which is the number of zeros of f counted with multiplicity; and
- $\deg D_2$, which is the number of poles of f counted with multiplicity.

Remark 32.16. Every principal divisor is of degree 0: that is, $\deg(\operatorname{div} f)=0$ for every $f\in\kappa(C)^\times$.

(We will not prove these last results, but you proved the last fact for $C=\mathbb{P}^1_k$ on your last homework assignment.)

32.4. **Linear equivalence and the Picard group.**

Definition 32.17. Divisors D_1 and D_2 are called linearly equivalent if there exists $f \in \kappa(C)^\times$ such that $D_1 - D_2 = \operatorname{div}(f)$. (Write $D_1 \sim D_2$ in this case.)

Linear equivalence is an equivalence relation. Each equivalence class $[D]$ is called a divisor class. Because the set of principal divisors is a subgroup of $\operatorname{Div} C$, the set of equivalence classes is the quotient group

$$\operatorname{Pic} C := \frac{\operatorname{Div} C}{\{\text{principal divisors}\}},$$

which is called the Picard group of C. Since $\operatorname{Div} C$ is abelian, so is its quotient $\operatorname{Pic} C$.

Example 32.18. Let $C = \mathbb{P}^1_k$. Two divisors on C are linearly equivalent if and only if they have the same degree. In other words, $\operatorname{Pic} C \simeq \mathbb{Z}$. (You proved this in your last homework assignment.)

In general, for any nice curve C over k, there is an exact sequence

$$0 \to k^\times \to \kappa(C)^\times \to \operatorname{Div} C \to \operatorname{Pic} C \to 0.$$

Remark 32.19. In more advanced algebraic geometry courses, one shows that divisor classes are in bijection with isomorphism classes of line bundles, which, loosely speaking, are families of vector spaces in which one has one vector space for each point of C.

Because principal divisors are of degree 0, the degree homomorphism

$$\begin{aligned} \operatorname{Div} C &\to \mathbb{Z} \\ D &\mapsto \deg D \end{aligned}$$

factors through the quotient $\operatorname{Pic} C$: i.e., it induces a well-defined homomorphism

$$\begin{aligned} \operatorname{Pic} C &\to \mathbb{Z} \\ [D] &\mapsto \deg D. \end{aligned}$$

Its kernel, consisting of divisor class of degree 0, is denoted $\operatorname{Pic}^0 C$.

Example 32.20. Let E be the projective closure of the affine curve E_0 in $\mathbb{A}^2_{\mathbb{Q}}$ given by

$$y^2 = x(x-1)(x-7).$$

We will show that $\operatorname{Pic} E$ contains an element of order 2.

The projective closure is given by the equation

$$y^2 z = x(x-z)(x-7z)$$

in $\mathbb{P}^2_{\mathbb{Q}}$. If we intersect with the "hyperplane at infinity" $z = 0$ in $\mathbb{P}^2$, we find that $x = 0$ too, so the point $\infty := (0:1:0)$ is the unique point on E not contained in the affine patch E_0.

What is the divisor of the rational function

$$x \in \kappa(E_0)^\times = \kappa(E)^\times?$$

On E_0, the function x vanishes only at $P := (0,0)$. But (x) must have total degree 0, so $(x) = nP - n\infty$ for some positive integer n. To find n, we compute $v_P(x)$. Since $\frac{\partial}{\partial x}(y^2 - x(x-1)(x-7)) \neq 0$ at P, the function y is a uniformizer at P. Then

$$x = \frac{1}{(x-1)(x-7)} y^2,$$

and the first factor is a unit at P, so $v_P(x) = 2$. Thus

$$(x) = 2P - 2\infty.$$

Let $D = P - \infty$. Then $2D$ is principal, so $[D] \in \operatorname{Pic} E$ satisfies $2[D] = 0$ in $\operatorname{Pic} E$.

How do we show that $[D]$ itself is not 0 in $\operatorname{Pic} E$? In other words, how do we show that D is not principal?

One way: if $D = (f)$, then $\deg f = 1$ (the number of zeros of f), so $[\kappa(E) : \kappa(\mathbb{P}^1)] = 1$, so E is birational to $\mathbb{P}^1$, which means that $E \simeq \mathbb{P}^1$ (since E and $\mathbb{P}^1$ are nice curves). But $E(\mathbb{R})$ has two connected components, and $\mathbb{P}^1(\mathbb{R})$ has only 1!

Another way: For simplicity, we can base extend to $\mathbb{C}$ (if $D = (f)$ for some $f \in \kappa(E)^\times$, then D viewed in $\operatorname{Div} E_{\mathbb{C}}$ is principal too, the divisor of the same f). Redefine E as the base extension $E_{\mathbb{C}}$. The function field $\kappa(E)$ is $\mathbb{C}(x)(\sqrt{x(x-1)(x-7)})$, which is a quadratic extension of $\mathbb{C}(x)$. The nontrivial automorphism σ (of order 2) of $\kappa(E)$ fixing $\mathbb{C}(x)$ induces an automorphism $\sigma\colon E \to E$ whose restriction to E_0 is the morphism $(x,y) \mapsto (x,-y)$. This σ induces an automorphism of $\operatorname{Div} E$, namely $\sum n_P P \mapsto \sum n_P({}^\Sigma P)$. In particular, if $(f) = D$, then

$${}^\sigma(f) = {}^\sigma D = {}^\sigma P - {}^\sigma\infty = P - \infty = D = (f),$$

which implies that ${}^\sigma f = cf$ for some $c \in k^\times$. Applying σ shows that $f = c\,{}^\sigma f$. Thus $f = c(cf) = c^2 f$, so $c = \pm 1$. If $c = 1$, then f is in the fixed field $\kappa(E)^\sigma$, so $f \in \mathbb{C}(x)$. If $c = -1$, then f/y is in the fixed field, so $f = g(x)y$ for some rational function $g \in \mathbb{C}(x)^\times$. For each $a \in \mathbb{C}$, the divisor of $x - a$ is

$$(a, \sqrt{a(a-1)(a-7)}) + (a, -\sqrt{a(a-1)(a-7)}) - 2\infty$$

(if $a \notin \{0,1,7\}$ then $x - a$ is a uniformizer at each of the first two points and has no other zeros or poles except at ∞; if $a \in \{0,1,7\}$ use an argument as for x above). And taking the divisors of both sides of the equation

$$y^2 = x(x-1)(x-7)$$

shows that

$$(y) = (0,0) + (1,0) + (7,0) - 3\infty.$$

The rational function f in $\mathbb{C}(x)^\times$ or $\mathbb{C}(x)^\times y$ is a product of these times a nonzero constant, so in the principal divisor (f), the multiplicities of $(0,0)$ and $(1,0)$ have the same parity. In particular, $D - \infty$ cannot equal (f).

33. Genus

(In this section you will be asked to take many things on faith, even more than usual.)

If C is a nice curve over $\mathbb{C}$, then $C(\mathbb{C})$ can be viewed as a topological space, and it turns out to be a 1-dimensional compact complex manifold; i.e., a compact Riemann surface. By the classification of compact oriented surfaces, it is homeomorphic to a sphere with g handles, i.e., a g-holed torus, for some nonnegative integer g. This integer g is called the genus of C.

In fact, there also exist algebraic definitions of the genus, in terms of "differentials", and these apply to nice curves over any field, even fields of characteristic p.

The genus of a nice curve is unchanged by base extension.

33.1. Newton polygons of two-variable polynomials.

Definition 33.1. A lattice point in the plane $\mathbb{R}^2$ is an element of $\mathbb{Z}^2$.

Definition 33.2. A convex lattice polygon P in $\mathbb{R}^2$ is the convex hull of a finite subset of $\mathbb{Z}^2$. (Loosely speaking, you put a rubber band around the points.) We (re)define the length of a side of P as $n-1$, where n is the number of lattice points on the side including the endpoints.

Suppose that C is a nice curve birational to an affine plane curve $f(x,y) = 0$, where

$$f(x,y) = \sum_{i,j} a_{ij} x^i y^j \in k[x,y].$$

Let P be a convex lattice polygon containing $\{(i,j) \in \mathbb{Z}^2 : a_{ij} \neq 0\}$. For instance P could be the Newton polygon of f, defined as the convex hull of $\{(i,j) \in \mathbb{Z}^2 : a_{ij} \neq 0\}$.

Given a side s of P, choose a direction along it, and label its lattice points $0, 1, \ldots, \ell$, where ℓ is the length of s; now form the homogeneous polynomial $f_s(t,u)$ of degree ℓ whose $\ell+1$ coefficients are the coefficients of f corresponding to the lattice points on s in order (choose one of the two possible directions along s). We call f_s a side polynomial (this is not standard terminology).

Theorem 33.3. *Let $f = \sum a_{ij} x^i y^j$ and P be as above. Suppose that*

(i) *The affine curve $f(x,y) = 0$ is smooth.*

(ii) *For each side s of P, the side polynomial f_s is squarefree.*

Then the genus of C equals the number of lattice points in the interior of P.

Remark 33.4. The zero polynomial is not squarefree. Thus the condition on the side polynomials will be satisfied usually only if P is close to being the Newton polygon on f.

34. Riemann-Roch theorem

Definition 34.1. Given $D \in \operatorname{Div} C$, define

$$L(D) := \{f \in \kappa(C)^\times : (f) + D \geq 0\} \cup \{0\}.$$

Proposition 34.2. *The set $L(D)$ is a k-subspace of $\kappa(C)$.*

Proof. Suppose that $D = \sum n_P P$. To say that $f \in L(D)$ is to say that $v_P(f) \geq -n_P$ for all P. Each condition $v_P(f) \geq -n_P$ defines a set of f that contains 0 and is closed under addition and multiplication by constants in k,

so each condition defines a subspace V_P of $\kappa(C)$. Then $L(D) = \bigcap_P V_P$, so $L(D)$ is a subspace too. □

Example 34.3. If $D = 0$, then $L(D)$ is the set of $f \in \kappa(C)$ such that $(f) \geq 0$. But for nonzero f, the divisor (f) has degree 0, so $(f) \geq 0$ is possible only if $(f) = 0$, which holds when $f \in k^\times$. Thus $L(D) = k$.

Example 34.4. If $D = 2P$ for a closed point P, then $L(D)$ is the set of $f \in \kappa(C)$ with at most a double pole at P (i.e., a double pole, simple pole, or defined at P), and defined at all other closed points of C. If $D = 3P - 2Q$, for closed points P and Q, then $L(D)$ is the set of $f \in \kappa(C)$ with at most a triple pole at P, and with at least a double zero at Q.

If $D_1 \leq D_2$, then $L(D_1) \subseteq L(D_2)$.

Example 34.5. Let $C = \mathbb{P}^1 \supset \mathbb{A}^1 = \operatorname{Spec} k[t]$. Let $\infty \in \mathbb{P}^1(k)$ be the point outside this $\mathbb{A}^1$, so $v_\infty(t) = -1$, and more generally $v_\infty(p(t)) = -\deg p$ for any polynomial $p(t) \in k[t]$. Let $D = 3\infty$.

What is $L(3\infty)$? If $f = \frac{p(t)}{q(t)} \in L(3\infty)$, where $p(t)$ and $q(t)$ are nonzero relatively prime polynomials in $k(t)$, then $q(t)$ cannot have a zero at any closed point P of $\mathbb{A}^1$, because at any such zero we would get $v_P(f) < 0$, so $(f) + 3\infty$ would not be effective. Thus $q(t)$ is a constant, and we may assume $q(t) = 1$. Thus $f = p(t)$ is a *polynomial* in t. The condition $(f) + 3\infty \geq 0$ implies $v_\infty(f) \geq -3$, which says that $-\deg p(t) \geq -3$, so $\deg p(t) \leq 3$. Thus $L(3\infty)$ is the k-vector space of polynomials in $k[t]$ of degree at most 3. In particular, $L(3\infty)$ has basis $1, t, t^2, t^3$, so $\dim_k L(3\infty) = 4$.

Let $P \in \mathbb{A}^1(k)$ be the point where t takes the value 7. What is $L(3\infty - P)$? This is the subspace of $L(3\infty)$ consisting of polynomials that have at least a simple zero at P, or equivalently, that are divisible by $t-7$. Thus $L(3\infty - P) = \{(t-7)g(t) : g(t) \in k[t], \deg g(t) \leq 2\}$, which is a 3-dimensional k-vector space.

It turns out that $\dim_k L(D)$ is always finite.

Definition 34.6. For each $D \in \operatorname{Div} C$, define $\ell(D) := \dim_k L(D) \in \mathbb{Z}_{\geq 0}$.

Example 34.7. If $D = 0$, then $L(D) = k$, so $\ell(D) = 1$.

Proposition 34.8. *If* $\deg D < 0$*, then* $L(D) = \{0\}$ *and* $\ell(D) = 0$*.*

Proof. Suppose that $\deg D < 0$. If $f \in L(D) - \{0\}$, then $(f) + D \geq 0$. The divisor (f) has degree 0, so $(f) + D$ has the same negative degree as D. On the other hand, if $(f) + D \geq 0$, then $(f) + D$ has nonnegative degree. This contradiction shows that no such f exists. □

Proposition 34.9. *If* D *and* D' *are linearly equivalent, then* $\ell(D) = \ell(D')$*.*

Proof. Write $D = D' + (g)$ for some $g \in \kappa(C)^\times$. If $f \in L(D)$ is nonzero, then $(f) + D \geq 0$, so $(f) + D' + (g) \geq 0$, so $(fg) + D' \geq 0$, so $fg \in L(D')$. Thus multiplication-by-g maps $L(D)$ into $L(D')$, and does so injectively, since multiplication-by-g on $\kappa(C)$ is injective. Similarly, multiplication-by-g^{-1} maps $L(D')$ into $L(D)$. These maps define inverse isomorphisms of k-vector spaces between $L(D)$ and $L(D')$. In particular, their dimensions $\ell(D)$ and $\ell(D')$ are the same. □

Theorem 34.10 (Riemann-Roch)**.** *Let* C *be a nice curve of genus* g *over* k*. There exists a divisor class consisting of divisors* K *called canonical divisors such that*

$$\ell(D) - \ell(K - D) = \deg D + 1 - g$$

for all $D \in \operatorname{Div} C$*.*

The Riemann-Roch theorem is rather deep, so we will not prove it here. From now on, K denotes any fixed canonical divisor.

Corollary 34.11.

(i) $\ell(K) = g$.
(ii) $\deg K = 2g - 2$.
(iii) *If* $\deg D > 2g - 2$*, then* $\ell(D) = \deg D + 1 - g$.

Proof.

(i) Taking $D = 0$ in the Riemann-Roch theorem yields

$$1 - \ell(K) = 0 + 1 - g,$$

so $\ell(K) = g$.

(ii) Taking $D = K$ yields

$$g - 1 = \deg K + 1 - g$$

so $\deg K = 2g - 2$.

(iii) If $\deg D > 2g - 2$, then $\deg(K - D) < 0$ so $\ell(K - D) = 0$ by Proposition 34.8. So the Riemann-Roch theorem simplifies to

$$\ell(D) = \deg D + 1 - g.$$

□

Example 34.12. Let $C = \mathbb{P}^1 \supset \mathbb{A}^1 = \operatorname{Spec} k[t]$. For $d \geq 0$, we have

$$L(d\infty) = \{p(t) \in k[t] : \deg p(t) \leq d\},$$

so $\ell(d\infty) = d + 1$. On the other hand, when d is sufficiently large, then Corollary 34.11(c) implies that $\ell(d\infty) = d + 1 - g$. Thus $g = 0$. (This agrees with the fact that $\mathbb{P}^1(\mathbb{C})$ is topologically a sphere, which is of genus 0.) If $D \in \operatorname{Div} \mathbb{P}^1$ is of degree $d > -2$, then Corollary 34.11(c) implies that $\ell(D) = \deg D + 1$; alternatively, use that $D \sim d\infty$ to obtain $\ell(D) = \ell(d\infty) = d + 1$. To summarize, for any $D \in \operatorname{Div} \mathbb{P}^1$ of degree d, we have

$$\ell(D) = \begin{cases} 0 & \text{if } d < 0 \text{ (by Proposition 34.8)} \\ d+1 & \text{if } d \geq 0. \end{cases}$$

The same conclusion holds for any genus 0 curve C, by a similar argument.

Proposition 34.13. *If C is a nice curve of genus 0 over k, and $C(k)$ is nonempty, then $C \simeq \mathbb{P}^1_k$.*

Proof. Choose $P \in C(k)$. By Corollary 34.11(c), $\ell(P) = 1 + 1 = 2$, but $\ell(0) = 1$ as in Example 34.7, so there exists $f \in L(P) - L(0)$. Since $L(0) = k$, this means that f is a nonconstant function with a simple pole at P and no other poles. The number of poles of f is 1, so the degree of the morphism $C \to \mathbb{P}^1$ given by $(f : 1)$ equals 1. In other words, $C \to \mathbb{P}^1$ is a birational map, and hence an isomorphism. □

35. Weierstrass equations

From now on, k is a perfect field of characteristic not 2 or 3.

Definition 35.1. A (short) Weierstrass equation is a polynomial equation of the form

$$y^2 = x^3 + Ax + B$$

for some constants $A, B \in k$. (If $\operatorname{char} k$ were 2 or 3, we would instead consider long Weierstrass equations of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

but when $\operatorname{char} k \neq 2, 3$, we can complete the square in y to make $a_1 = 0$ and $a_3 = 0$, and then complete the cube in x to make $a_2 = 0$.)

Proposition 35.2. *Let E be the projective closure in $\mathbb{P}^2$ of the affine curve E_0 defined by a Weierstrass equation $y^2 = x^3 + Ax + B$. Then the following are equivalent:*

(i) *The affine curve E_0 is smooth.*
(ii) *The projective curve E is smooth.*

(iii) *The polynomial* $x^3 + Ax + B$ *is separable (or equivalently, squarefree).*
(iv) *The* discriminant $-16(4A^3 + 27B^2)$ *is nonzero.*

If these conditions hold, then E *is a nice genus* 1 *curve with a single point* $P := (0 : 1 : 0)$ *at infinity. Otherwise,* E *has a unique singular point, and* E *is birational to* $\mathbb{P}^1$*, which is of genus* 0*.*

Proof. This will be assigned for homework. If E is smooth, the fact that its genus is 1 follows from the genus formula $(d-1)(d-2)/2$ for a smooth plane curve of degree d. □

36. Elliptic curves

Definition 36.1. An elliptic curve over k is a nice genus-1 curve over k equipped with a point in $E(k)$ (called the origin).

Theorem 36.2.

(i) *Given a Weierstrass equation* $y^2 = x^3+Ax+B$ *with* x^3+Ax+B *separable, the projective closure of this affine curve, equipped with the point* $(0 : 1 : 0)$*, is an elliptic curve over* k*.*
(ii) *Every elliptic curve over* k *is isomorphic to one arising in this way.*

Proof.

(i) This follows from Proposition 35.2.
(ii) Let E be an elliptic curve, so E is of genus 1. Let $P \in E(k)$ be the origin of E. By Corollary 34.11(c), we have $\ell(nP) = n$ for all $n \geq 1$. So we have bases as follows:

$$\begin{aligned} L(0) &= \langle 1 \rangle = k, \\ L(P) &= \langle 1 \rangle, \\ L(2P) &= \langle 1, x \rangle \qquad \text{for some } x \in \kappa(E), \\ L(3P) &= \langle 1, x, y \rangle \qquad \text{for some } y \in \kappa(E). \end{aligned}$$

In particular, $v_P(x) = -2$ and $v_P(x) = -3$. Then $v_P(x^2) = -4$, so $x^2 \in L(4P) - L(3P)$. Thus

$$L(4P) = \langle 1, x, y, x^2 \rangle.$$

Similarly,

$$L(5P) = \langle 1, x, y, x^2, xy \rangle.$$

Now the 7 functions $1, x, y, x^2, xy, x^3, y^2$ in the 6-dimensional vector space $L(6P)$ must be linearly dependent, and the relation must involve both x^3 and y^2 since both of these have valuation -6 at P. By replacing x, y by $\lambda x, \lambda y$ for suitable $\lambda \in k^\times$, we may assume that the relation takes the form of a long Weierstrass equation. By completing the square and cube, we may make it a short Weierstrass equation instead. Let C be

the nice curve birational to the one given by this Weierstrass equation. Then $\kappa(C) = k(x,y) \subseteq \kappa(E)$. Since

$$[\kappa(E) : k(x)] = \deg x = \#\{\text{poles of } x, \text{ counted with multiplicity}\} = 2$$

and $[\kappa(E) : k(y)] = 3$ are relatively prime, $k(x,y) = \kappa(E)$. Thus E is birational to C. So C has genus 1, so $x^3 + Ax + B$ is separable by Proposition 35.2.

□

37. Group law

Theorem 37.1. *Let E be an elliptic curve with origin O. Then the map of sets*

$$E(k) \mapsto \operatorname{Pic}^0 E$$
$$P \mapsto [P - O]$$

is a bijection.

Proof. Injectivity: Suppose that $P, Q \in E(k)$ are distinct points such that $[P-O] = [Q-O]$. Then $P-Q = (f)$ for some $f \in \kappa(E)^\times$. This f is of degree 1, so $(f : 1)$ defines an isomorphism $E \to \mathbb{P}^1$, contradicting the assumption that E has genus 1.

Surjectivity: Let $[D] \in \operatorname{Pic}^0 E$, where $D \in \operatorname{Div}^0 E$. Then $\ell(D+O) = 1$, so $L(D+O) \neq \{0\}$, so there exists $f \in \kappa(E)^\times$ such that $(f) + D + O \geq 0$. But $\deg((f) + D + O) = 0 + 0 + 1 = 1$, so $(f) + D + O = P$ for some $P \in E(k)$. Thus $[D] = [P - O]$. □

Since $\operatorname{Pic}^0 E$ is an abelian group, the bijection above makes $E(k)$ into an abelian group.

37.1. **Chord-tangent description.** Let $E \subseteq \mathbb{P}^2$ be an elliptic curve in Weierstrass form. Let $L \subseteq \mathbb{P}^2$ be a line. Then $L \cap E$ can be computed by changing coordinates on $\mathbb{P}^2$ to make L given by $z = 0$, and then substituting $z = 0$ into the degree 3 homogeneous polynomial defining E to get a degree 3 homogeneous polynomial in $k[x,y]$, and looking at its zeros on $\mathbb{P}^1 \simeq L$. The result is three $\overline{k}$-points, if we count them with appropriate multiplicities. More precisely, we may view $L \cap E$ as a divisor of degree 3 on E.

Example 37.2. If L is the line at infinity, given by $z = 0$, then $L \cap E$ gives the divisor $3 \cdot O$ since substituting $z = 0$ into

$$y^2 z = x^3 + Axz^2 + Bz^3$$

yields $x^3 = 0$.

Let L_1 and L_2 be two lines in $\mathbb{P}^2$, defined by linear forms ℓ_1 and ℓ_2, respectively. View $f := \ell_1/\ell_2$ as a rational function on E. Then one can show that

$$(f) = (L_1 \cap E) - (L_2 \cap E)$$

where the intersections are viewed as degree 3 divisors on E as above. In particular, if L_2 is the line at infinity, and $L_1 \cap E = (P) + (Q) + (R)$, where $P, Q, R \in E(k)$, then $(f) = (P) + (Q) + (R) - 3 \cdot O = (P - O) + (Q - O) + (R - O)$.

Proposition 37.3. *Let $E \subseteq \mathbb{P}^2$ be an elliptic curve in Weierstrass form, and let $O = (0 : 1 : 0)$, as usual. Then*

(i) *The point O is the identity for the group law on $E(k)$.*

(ii) *If $P, Q, R \in E(k)$ are such that there is a line L with $L \cap E = (P) + (Q) + (R)$, then $P + Q + R = O$ in the group $E(k)$.*

Proof. (1) The point $O \in E(k)$ corresponds to $[O - O] \in \operatorname{Pic}^0 E$.

(2) The sum $P + Q + R$ in $E(k)$ corresponds to $[P - O] + [Q - O] + [R - O]$, which as explained just before this proposition, is the class of a principal divisor.

□

Proposition 37.3 characterizes the group law on $E(k)$ completely:

- To compute the inverse of a point $P = (a, b) \in E(k)$ not equal to O, let $L \subseteq \mathbb{P}^2$ be the projective closure of the vertical line $x = a$ in $\mathbb{A}^2$; then $L \cap E = (P) + (P') + (O)$, where $P' := (a, -b)$. (L passes through O since its homogeneous equation is $x - az = 0$, which vanishes at $(0 : 1 : 0)$); thus according to Proposition 37.3(ii), $P + P' + O = O$, so $P' = -P$. Of course, Proposition 37.3(i) we also know that $-O = O$.
- To compute $P + Q$ where $P, Q \in E(k)$, first let L be the line in $\mathbb{P}^2$ through P and Q; if $P = Q$, take L to be the tangent line to E at P. Then $L \cap E = (P) + (Q) + (R)$ for some $R \in E(k)$ (it is a k-point because its degree must be 1; more concretely, it is so because if two roots of a cubic polynomial are rational, then the third root is rational too). By Proposition 37.3(ii), $P + Q + R = O$, so $P + Q = -R$, which can be determined, as we already saw.

In fact, it is possible to define a product variety $E \times E$, an addition morphism $E \times E \to E$, and an inverse morphism $E \to E$.

37.2. **Torsion points.**

Definition 37.4. Let E be an elliptic curve over k. Let $P \in E(L)$ for some field extension $L \supseteq k$. Let $n \in \mathbb{Z}_{\geq 1}$. Call P an *n-torsion point* if $nP = O$ in the group $E(L)$. The *n-torsion subgroup* $E[n]$ of $E(\overline{k})$ is the kernel of the multiplication-by-n homomorphism

$$\begin{aligned} [n] \colon E(\overline{k}) &\to E(\overline{k}) \\ P &\mapsto nP. \end{aligned}$$

Example 37.5. Assume $\operatorname{char} k \neq 2$. Let E be the projective closure of $y^2 = f(x)$ where $f(x)$ is a separable cubic polynomial. Then $E[2]$ consists of O and the three points $(\alpha, 0)$ where $\alpha \in \overline{k}$ is a zero of f. Thus $E[2] \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Assume moreover that $f(x) = (x-e_1)(x-e_2)(x-e_3)$ for distinct $e_1, e_2, e_3 \in k$. Then $E[2] \subseteq E(k)$. Consider the multiplication-by-2 morphism on E and the corresponding extension of function fields.

$$\begin{array}{ccc} E & & L := \kappa(E) \\ \Big\downarrow{\scriptstyle [2]} & & \Big\uparrow \\ E & & K := \kappa(E) \end{array}$$

Each $T \in E[2]$ induces an addition-by-T morphism $\tau_T \colon E \to E$, a deck transformation of the top E (i.e., an automorphism satisfying $[2](\tau_T(P)) = [2](P)$), and this corresponds to an automorphism of L acting trivially on K. In fact, we get an injective homomorphism $E[2] \to \mathrm{Aut}(L/K)$.

In fact, it turns out that $[2] \colon E \to E$ is a morphism of degree $2^2 = 4$ (that is, $[L : K] = 4$), so L/K is a Galois extension with Galois group $E[2]$. One way to prove this is to compute degrees of all the morphisms in the diagram

$$\begin{array}{ccc} E & \xrightarrow{[2]} & E \\ \Big\downarrow{\scriptstyle x} & & \Big\downarrow{\scriptstyle x} \\ \mathbb{P}^1 & \xrightarrow{\phi} & \mathbb{P}^1 \end{array}$$

where x is the projection onto the x-coordinate, and $\phi(x)$ is the rational function giving $x([2]P)$ for $P = (x, y)$: an explicit calculation of the tangent line for $y^2 = x^3 + Ax + B$ gives

$$x \mapsto \phi(x) := x([2]P) = \frac{x^4 - 2Ax^2 - 8Bx + A^2}{4(x^3 + Ax + B)},$$

which is a rational function of degree $\max(4, 3) = 4$. For a more conceptual proof that $\deg[2] = 4$, using differentials and dual isogenies, see [Sil92].

Since $\mathrm{Gal}(L/K) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, there are three intermediate quadratic fields, and it turns out that these are $K(\sqrt{x - e_i})$ for $i = 1, 2, 3$. Note that $(x - e_1)(x - e_2)(x - e_3)$ is already a square in K, namely y^2. So $L = K(\sqrt{x - e_1}, \sqrt{x - e_2})$.

38. Mordell's theorem

In a 1901 paper, Poincaré considered the problem of finding generators for the group $E(\mathbb{Q})$ for an elliptic curve E over $\mathbb{Q}$. It was only many years later, in 1922, that Mordell proved the existence of a finite set of generators. He used an argument resembling the "method of infinite descent" used by Fermat to prove that $x^4 + y^4 = z^2$ has no solutions in positive integers.

Theorem 38.1 (Mordell)**.** *If E is an elliptic curve over $\mathbb{Q}$, then the abelian group $E(\mathbb{Q})$ is finitely generated.*

By the structure theorem for finitely generated abelian groups, Mordell's theorem implies that $E(\mathbb{Q}) \simeq \mathbb{Z}^r \times T$ for some nonnegative integer r (called the rank of E) and some finite abelian group T (called the torsion subgroup of E).

Remark 38.2. Mordell's theorem is sometimes also called the Mordell-Weil theorem, but Weil's contribution was to generalize it by replacing $\mathbb{Q}$ with an arbitrary finite extension of $\mathbb{Q}$ and E by an abelian variety of arbitrary dimension.

All known proofs of Theorem 38.1 are minor variants of the one we will give. It consists of two parts. The first part is the following:

Theorem 38.3 (Weak Mordell-Weil theorem)**.** *If E is an elliptic curve over $\mathbb{Q}$, then $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.*

The second part involves the construction of a function $h\colon E(\mathbb{Q}) \to \mathbb{R}$ called a height function. For $P \in E(\mathbb{Q})$, the value $h(P)$ measures the size of the numerators and denominators of the coordinates of P.

Remark 38.4. It is not known whether there exists an algorithm that takes E as input and outputs a finite list of points that generate $E(\mathbb{Q})$. The problem is that the proof of the weak Mordell-Weil theorem is not effective; i.e., it does not produce coset representative for the elements of $E(\mathbb{Q})/2E(\mathbb{Q})$, even in principle.

39. The weak Mordell-Weil theorem

In this section we will prove the weak Mordell-Weil theorem in the case that $E[2] \subseteq E(\mathbb{Q})$, i.e., the case in which E is given by an equation of the form

$$y^2 = (x - e_1)(x - e_2)(x - e_3).$$

If we make the substitution $x = x'/d^2$ and $y = y'/d^3$ and multiply both sides by d^6, we get an isomorphic curve; moreover, by choosing d so the denominator of each e_i divides d, the new curve is of the same form but with $e_i \in \mathbb{Z}$. *So assume that $e_i \in \mathbb{Z}$ from now on.*

Lemma 39.1. *We have an isomorphism of abelian groups*

$$\begin{aligned} \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}} &\xrightarrow{\sim} \mathrm{Hom}_{\mathrm{conts}}(G_{\mathbb{Q}}, \{\pm 1\}) \\ \bar{a} &\mapsto \left(\sigma \mapsto \frac{{}^\sigma\sqrt{a}}{\sqrt{a}}\right). \end{aligned}$$

(Here, for each $a \in \mathbb{Q}^\times$, we write $\bar{a}$ for its image in $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$, and $\sqrt{a}$ for a fixed square root of a in $\overline{\mathbb{Q}}^\times$. The notation $\mathrm{Hom}_{\mathrm{conts}}$ denotes the group of continuous homomorphisms.)

Proof. First let us show that $\sigma \mapsto \frac{{}^{\sigma}\sqrt{a}}{\sqrt{a}}$ is a homomorphism: If $\sigma, \tau \in G_{\mathbb{Q}}$, then

$$\begin{aligned}\frac{{}^{\sigma\tau}\sqrt{a}}{\sqrt{a}} &= \frac{{}^{\sigma}\sqrt{a}}{\sqrt{a}} \cdot {}^{\sigma}\left(\frac{{}^{\tau}\sqrt{a}}{\sqrt{a}}\right) \\ &= \frac{{}^{\sigma}\sqrt{a}}{\sqrt{a}} \cdot \frac{{}^{\tau}\sqrt{a}}{\sqrt{a}},\end{aligned}$$

since the number $\frac{{}^{\tau}\sqrt{a}}{\sqrt{a}} = \pm 1$ is fixed by σ. It is a continuous homomorphism, since its kernel is the closed subgroup $G_{\mathbb{Q}(\sqrt{a})}$ of $G_{\mathbb{Q}}$. Next, this homomorphism is independent of the choice of $\sqrt{a}$, since changing the sign of $\sqrt{a}$ does not change the ratio $\frac{{}^{\sigma}\sqrt{a}}{\sqrt{a}}$. Thus we have a well defined map of sets

$$\mathbb{Q}^{\times} \xrightarrow{\partial} \mathrm{Hom}_{\mathrm{conts}}(G_{\mathbb{Q}}, \{\pm 1\}).$$

If $a, b \in \mathbb{Q}^{\times}$, and we choose square roots $\sqrt{a}$ and $\sqrt{b}$, and use $\sqrt{a} \cdot \sqrt{b}$ as a square root of ab, we get

$$\frac{{}^{\sigma}\sqrt{ab}}{\sqrt{ab}} = \frac{{}^{\sigma}\sqrt{a}}{\sqrt{a}} \frac{{}^{\sigma}\sqrt{b}}{\sqrt{b}},$$

so ∂ is a homomorphism. We have

$$\begin{aligned} a \in \ker(\partial) &\iff {}^{\sigma}\sqrt{a} = \sqrt{a} \qquad \text{for all } \sigma \in G_{\mathbb{Q}} \\ &\iff \sqrt{a} \in \mathbb{Q}^{\times} \\ &\iff a \in \mathbb{Q}^{\times 2}. \end{aligned}$$

Thus ∂ induces a homomorphism

$$\frac{\mathbb{Q}^{\times}}{\mathbb{Q}^{\times 2}} \xrightarrow{\partial} \mathrm{Hom}_{\mathrm{conts}}(G_{\mathbb{Q}}, \{\pm 1\}).$$

Given a nontrivial element $\bar{a} \in \frac{\mathbb{Q}^{\times}}{\mathbb{Q}^{\times 2}}$, we get a well-defined quadratic extension $L := \mathbb{Q}(\sqrt{a})$. Given a quadratic extension L of $\mathbb{Q}$, we get a nontrivial continuous homomorphism $G_{\mathbb{Q}} \to \mathrm{Gal}(L/\mathbb{Q}) \simeq \{\pm 1\}$. The composition of these constructions is ∂. Moreover, each construction can be reversed: Given a nontrivial continuous homomorphism $G_{\mathbb{Q}} \simeq \{\pm 1\}$, its kernel is a closed subgroup of index 2 in $G_{\mathbb{Q}}$, which by Galois theory is G_L for some field L of degree 2 over $\mathbb{Q}$. And given a quadratic extension L of $\mathbb{Q}$, we may write $L = \mathbb{Q}(\sqrt{a})$ where $a \in \mathbb{Q}^{\times}$ is uniquely determined modulo squares. This completes the proof that our homomorphism is an isomorphism. □

There is a partial analogue in which the multiplicative group of a field is replaced by an elliptic curve:

Lemma 39.2. *For any elliptic curve E over $\mathbb{Q}$ such that $E[2] \subseteq E(\mathbb{Q})$, there is an injective homomorphism*

$$\begin{aligned} \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} &\hookrightarrow \mathrm{Hom}_{\mathrm{conts}}(G_{\mathbb{Q}}, E[2]) \\ \bar{P} &\mapsto (\sigma \mapsto {}^{\sigma}Q - Q)\,. \end{aligned}$$

(Here, for each $P \in E(Q)$, we write $\bar{P}$ for its image in $E(\mathbb{Q})/2E(\mathbb{Q})$, and Q for a point in $E(\overline{\mathbb{Q}})$ such that $2Q = P$.)

Proof. By Remark 31.5, a choice of Q exists for each P. If $\sigma \in G_{\mathbb{Q}}$, then $2 \cdot {}^{\sigma}Q = {}^{\sigma}(2Q) = {}^{\sigma}P = P$, so $2({}^{\sigma}Q - Q) = P - P = O$, so ${}^{\sigma}Q - Q \in E[2]$. The proof that $\sigma \mapsto {}^{\sigma}Q - Q$ is a homomorphism is the same as in Lemma 39.1; it is here that we use that $E[2]$ is fixed pointwise by every $\sigma \in G_{\mathbb{Q}}$. The rest of the proof also copies the proof of Lemma 39.1. □

Remark 39.3. There is a generalization of Lemma 39.2 that works even if $E[2]$ is not contained in $E(\mathbb{Q})$. It involves replacing $\mathrm{Hom}_{\mathrm{conts}}(G_{\mathbb{Q}}, E[2])$ by a continuous cohomology group, $\mathrm{H}^1(G_{\mathbb{Q}}, E[2])$.

Proposition 39.4. *For any elliptic curve E over $\mathbb{Q}$ such that $E[2] \subseteq E(\mathbb{Q})$, there is an injective homomorphism*

$$\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \overset{\phi}{\hookrightarrow} \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}} \times \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}}$$

If $P = (x, y) \in E(\mathbb{Q}) - \{O, (e_1, 0), (e_2, 0)\}$, then

$$\phi(\bar{P}) = (x - e_1, x - e_2)\,.$$

Also, $\phi(\bar{O}) = (1, 1)$ and

$$\begin{aligned} \phi(\overline{(e_1, 0)}) &= ((e_1 - e_2)(e_1 - e_3), e_1 - e_2) \\ \phi(\overline{(e_2, 0)}) &= (e_2 - e_1, (e_2 - e_1)(e_2 - e_3))\,. \end{aligned}$$

Remark 39.5. More canonically, we have an injective homomorphism

$$\begin{aligned} \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} &\hookrightarrow \ker\left(\left(\frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}}\right)^{\oplus 3} \to \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}}\right) \\ \overline{(x, y)} &\mapsto (x - e_1, x - e_2, x - e_3) \qquad (\text{for } (x, y) \in E(\mathbb{Q}) - E[2]). \end{aligned}$$

For $(x, y) \in E[2] - \{O\}$, two of the $x - e_i$ make sense, and the third can be assigned the value such that the product is 1. This explains the last formulas in Proposition 39.4.

Sketch of proof. The fact that this defines a homomorphism can be checked with a brute force calculation.

But it really comes from Lemma 39.2, plus the isomorphism

$$\begin{aligned} E[2] &\to \{\pm 1\} \times \{\pm 1\} \\ (e_1, 0) &\mapsto (-1, 1) \\ (e_2, 0) &\mapsto (1, -1), \end{aligned}$$

plus Lemma 39.1. It is saying that in order to take half of a point $(x, y) \in E(\mathbb{Q}) - E[2]$, one must adjoin $\sqrt{x - e_1}$ and $\sqrt{x - e_2}$ to the ground field. □

Final exam on Mon Dec 14, 9am-12 in 3-135. It will be mainly based on topics covered in homework problems. Remaining office hours this week: Wed 1:30-2:30, Fri *12:30-1:30*.

Challenge problems: Show that every nice genus 2 curve over a field of characteristic not 2 is birational to an affine curve $y^2 = f(x)$ with $f(x)$ separable of degree 5 or 6.

What can you say about explicit equations of genus 3 curves?

Compute $E(\mathbb{Q})/2E(\mathbb{Q})$ for $E\colon y^2 = x^3 - x$. Can you determine $E(\mathbb{Q})$ itself?

Proposition 39.6. *Let S be the set of primes p such that $p|(e_i - e_j)$ for some distinct i, j. Let $\mathbb{Q}(S, 2)$ be the finite subgroup of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ generated by (the images of) -1 and the primes in S. Then the image of the injective homomorphism*

$$\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \xrightarrow{\phi} \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}} \times \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}}$$

is contained in $\mathbb{Q}(S, 2) \times \mathbb{Q}(S, 2)$.

Sketch of proof. Suppose $P = (x, y) \in E(\mathbb{Q})$. For simplicity, let us assume that $P \notin E[2]$. To say that $x - e_1 \in \mathbb{Q}(S, 2)$ is to say that $v_p(x - e_1)$ is even for every prime $p \notin S$. Fix $p \notin S$.

Case 1: $v_p(x) < 0$. Then $v_p(x - e_i) = v_p(x)$ for $i = 1, 2, 3$. Now

$$\begin{aligned} 2v_p(y) &= v_p(y^2) \\ &= v_p((x - e_1)(x - e_2)(x - e_3)) \\ &= v_p(x - e_1) + v_p(x - e_2) + v_p(x - e_3) \\ &= 3v_p(x), \end{aligned}$$

so $v_p(x)$ is even.

Case 2: $v_p(x) \geq 0$. Then p divides at most one of $x - e_1$, $x - e_2$, $x - e_3$, because otherwise subtracting would show that p divides some $e_i - e_j$, so $p \in S$, a contradiction. On the other hand, $v_p((x - e_1)(x - e_2)(x - e_3))$ is even, as in Case 1, so $v_p(x - e_i)$ must be even *for each* i. □

Proposition 39.6 proves that $E(\mathbb{Q})/2E(\mathbb{Q})$ injects into a finite group; this proves the weak Mordell-Weil theorem (at least for elliptic curves E over $\mathbb{Q}$ with $E[2] \subset E(\mathbb{Q})$).

40. Height of a rational number

Definition 40.1. Let $t = a/b$ be a rational number in lowest terms. The (exponential) height of t is

$$H(t) := \max(|a|, |b|).$$

Extend the definition to $t \in \mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$ by defining $H(\infty) = 1$.

Definition 40.2. The (logarithmic) height of $t \in \mathbb{Q} \cup \{\infty\}$ is

$$h(t) := \log H(t).$$

Example 40.3. We have $h(100) < h(1001/1000)$.

In general, $h(t)$ is approximately the width of a piece of paper needed to write down t explicitly as a fraction of integers.

Proposition 40.4 (Northcott)**.** *For any bound $B \in \mathbb{R}$, the set $\{t \in \mathbb{Q} : H(t) \le B\}$ is finite.*

Proof. Each t in this set has numerator in the range $[-B, B]$ and denominator in $[1, B]$, so there are at most $(2B+1)B$ possibilities. □

Challenge problem: Find an asymptotic formula for the size of this set as $B \to \infty$.

Definition 40.5. The degree of a rational function $p(x)/q(x) \in \mathbb{Q}(x)$ in lowest terms is $\max(\deg p, \deg q)$.

Theorem 40.6. *If $f(x)$ is a rational function of degree d, then $h(f(t)) = dh(t) + O_f(1)$ for all $t \in \mathbb{Q}$. (That is, there is a constant $C = C(f)$ such that $|h(f(t)) - dh(t)|$ for all $t \in \mathbb{Q}$.)*

Proof. Write $f(x) = p(x)/q(x)$, where $p(x), q(x) \in \mathbb{Z}[x]$ have gcd 1.

Upper bound: Write $t = a/b$ in lowest terms. Let $P(x, y) = y^d\, p(x/y)$ and $Q(x, y) = y^d\, q(x/y)$ be the homogenizations of $p(x)$ and $q(x)$, respectively. Then $f(t) = f(a/b) = P(a, b)/Q(a, b)$. This might not be in lowest terms, but in any case

$$H(f(t)) \le \max(|P(a, b)|, |Q(a, b)|) \le C \max(|a|, |b|)^d = CH(t)^d$$

for some constant C depending on P and Q (i.e., on f). Taking log of both sides yields

$$h(f(t) \le dh(t) + \log C.$$

Lower bound: We must bound $|a|$ and $|b|$ in terms of $|P(a, b)|$ and $|Q(a, b)|$. Example: If $P(a, b) = 3a^2 + b^2$ and $Q(a, b) = ab$, we could use the identities

$$aP(a, b) - bQ(a, b) = 3a^3$$
$$bP(a, b) - 3aQ(a, b) = b^3.$$

In particular,

$$\gcd(P(a, b), Q(a, b))| \gcd(3a^3, b^3)|3$$

so $P(a,b)/Q(a,b)$ is almost in lowest terms, so

$$H(f(t)) = H(P(a,b)/Q(a,b)) \sim \max(|P(a,b)|, |Q(a,b)|),$$

where $\sim$ means up to a bounded constant factor. Also,

$$\begin{aligned} 3|a|^3 &\leq \max(|a|,|b|)\max(|P(a,b)|,|Q(a,b)|) \\ |b|^3 &\leq \max(|a|,|b|)\max(|P(a,b)|,|Q(a,b)|), \end{aligned}$$

so

$$\begin{aligned} \max(|a|,|b|)^3 &\leq \max(|a|,|b|)\max(|P(a,b)|,|Q(a,b)|) \\ \max(|a|,|b|)^2 &\leq \max(|P(a,b)|,|Q(a,b)|) \\ H(t)^2 &\leq H(f(t)) \text{ times a constant.} \\ 2h(t) &\leq h(f(t)) + O(1). \end{aligned}$$

To generalize to arbitrary $P(a,b)$ and $Q(a,b)$, we need the two identities. Observe that $P(a,b)$ and $Q(a,b)$ have no common zeros in $\overline{\mathbb{Q}}$ except $(0,0)$. So the Nullstellensatz implies that the ideals $(P(a,b),Q(a,b))$ and (a,b) of $\mathbb{Q}[a,b]$ have the same radical. In particular, for some n, we have that a^n and b^n lie in the ideal generated by $P(a,b)$ and $Q(a,b)$ in $\mathbb{Q}[a,b]$. Clearing denominators shows that there exists $c \in \mathbb{Z}_{\geq 1}$ such that the same holds for ca^n and cb^n in $\mathbb{Z}[a,b]$. □

41. Height functions on elliptic curves

Recall that we are studying the elliptic curve with equation

$$y^2 = (x-e_1)(x-e_2)(x-e_3).$$

Without loss of generality, by making the substitution $x \mapsto x+c$ for some $c \in \mathbb{Q}$, we may assume that the coefficient of x^2 in the right hand side is 0. And then, as before, we may also assume that $e_i \in \mathbb{Z}$ for all i. Now the right hand side is also $x^3 + Ax + B$ for some A and B.

Definition 41.1. For $P \in E(\mathbb{Q})$, define

$$h_x(P) := h(x(P)) = \log H(x(P)).$$

(By convention, $h_x(O) = 0$.)

Proposition 41.2. *For all $P \in E(\mathbb{Q})$, we have*

$$h_x(2P) = 4h_x(P) + O_E(1)$$

where the bound on the error term depends only on E, not on P.

Proof. We claim that there is a rational function $r(x)$ of degree 4 such that if $P = (x, y)$, then $x(2P) = r(x)$. This can be deduced by coordinate geometry, by using the chord-tangent law: one gets

$$x(2P) = \frac{x^4 - 2Ax^2 - 8Bx + A^2}{4(x^3 + Ax + B)}.$$

Alternatively, the diagram of curves

$$\begin{array}{ccc} E & \xrightarrow{[2]} & E \\ \downarrow{\scriptstyle x} & & \downarrow{\scriptstyle x} \\ \mathbb{P}^1 & \longrightarrow & \mathbb{P}^1, \end{array}$$

induces a diagram of function fields

$$\begin{array}{ccc} \mathbb{Q}(x,y) & \longleftarrow & \mathbb{Q}(x_2,y_2) \\ \uparrow & & \uparrow \\ \mathbb{Q}(x) & \longleftarrow & \mathbb{Q}(x_2), \end{array}$$

in which the coordinate functions x_2 and y_2 on the E on the right pull back to the functions $x(2P)$ and $y(2P)$ in $\mathbb{Q}(x, y)$. Since $[2]\colon E \to E$ is of degree $2^2 = 4$, computing degrees of all field extensions in the diagram shows that $2 \deg r = 4 \cdot 2$, so $\deg r = 4$. This completes the second proof of the claim.

Now, taking the height of both sides of $x(2P) = r(x)$ yields

$$h_x(2P) = h(r(x)) = 4h(x) + O_E(1),$$

by Theorem 40.6. (The function $r(x)$ depends on E, so the $O(1)$ depends on E too.) □

Lemma 41.3. *Given that E has equation $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$, Every rational point on E other than O has the form $\left(\frac{a}{d^2}, \frac{b}{d^3}\right)$ for some $a, b, d \in \mathbb{Z}$ with $\gcd(a, d) = 1$ and $\gcd(b, d) = 1$.*

Proof. Any $(x, y) \in E(\mathbb{Q}) - \{O\}$ satisfies

$$y^2 = x^3 + Ax + B.$$

Taking denominators of both sides shows that

$$\operatorname{denom}(y)^2 = \operatorname{denom}(x)^3,$$

where $\operatorname{denom}(x)$ denotes the positive integer denominator when x is written in lowest terms. (Another way to see this: The equation implies that $v_p(y) < 0$ if and only if $v_p(x) < 0$, and in that case, $2v_p(y) = 3v_p(x)$.) This implies that there exists $d \geq 1$ such that

$$\operatorname{denom}(y) = d^3 \qquad \text{and} \qquad \operatorname{denom}(x) = d^2.$$

□

Lemma 41.4. *If $P, Q \in E(\mathbb{Q}) - \{O\}$ satisfy $x(P) \neq x(Q)$, then*

$$x(P) + x(Q) + x(P+Q) = \left(\frac{y(Q) - y(P)}{x(Q) - x(P)}\right)^2.$$

Proof. Let $y = mx + b$ be the line through P and Q, so

$$m = \left(\frac{y(Q) - y(P)}{x(Q) - x(P)}\right).$$

That line intersects E in three points: P, Q, and R, say. Then $R = -(P+Q)$. Also, $x(P)$, $x(Q)$, $x(R)$ are the solutions to the cubic equation

$$(mx + b)^2 = x^3 + Ax + B,$$

or equivalently,

$$x^3 - m^2x^2 + (A - 2mb)x + (B - b^2) = 0,$$

so $x(P) + x(Q) + x(R) = m^2$. Substitute the value of m, and observe that $x(R) = x(P+Q)$ since $R = -(P+Q)$. □

Proposition 41.5. *Fix $P_0 \in E(\mathbb{Q})$. Then for every $P \in E(\mathbb{Q})$,*

$$h_x(P + P_0) \leq 2h_x(P) + O_{E,P_0}(1).$$

Proof. We may assume that $P_0 \neq O$. By increasing the constant, we can ignore any finite set of P, and hence assume that P is not O or $\pm P_0$. Write

$$P = (x, y) = \left(\frac{a}{d^2}, \frac{b}{d^3}\right)$$

as in Lemma 41.3. Similarly, write

$$P_0 = (x_0, y_0) = \left(\frac{a_0}{d_0^2}, \frac{b_0}{d_0^3}\right)$$

Then

$$x(P + P_0) = \left(\frac{y - y_0}{x - x_0}\right)^2 - x - x_0.$$

If we expand the square, replace y^2 by $x^3 + Ax + B$, and replace y_0^2 by $x_0^3 + Ax_0 + B$, then we eventually get

$$\begin{aligned} x(P + P_0) &= \frac{(xx_0 + A)(x + x_0) + 2B - 2yy_0}{(x - x_0)^2} \\ &= \frac{(aa_0 + Ad^2d_0^2)(ad_0^2 + a_0d^2) + 2Bd^4d_0^4 - 2bdb_0d_0}{(ad_0^2 - a_0d^2)^2}. \end{aligned}$$

Examining each monomial in the numerator and denominator shows that

$$H(x(P + P_0)) = O_{E,P_0}(1) \max\left\{|a|^2, |ad^2|, |d|^4, |bd|\right\}.$$

We have

$$|a| \leq H(x)$$
$$|d^2| \leq H(x)$$

and the equation $y^2 = x^3 + Ax + B$ so $b^2 = a^3 + Aad^4 + Bd^6$, so

$$|b|^2 \leq O_{E,P_0}(1)H(x)^3.$$

Plugging these estimates in yields

$$H(x(P + P_0)) = O_{E,P_0}(1)H(x)^2.$$

Taking log of both sides gives

$$h_x(P + P_0) \leq 2h_x(P) + O_{E,P_0}(1).$$

□

42. Descent

It is traditional to define the naïve height on the abelian group $G := E(\mathbb{Q})$ by the formula

$$h(P) = \frac{1}{2}h_x(P).$$

By Propositions 41.5, 41.2, and 40.4, respectively, $h\colon G \to \mathbb{R}$ satisfies the following axioms:

(i) For each $P_0 \in G$, we have $h(P + P_0) \leq 2h(P) + O_{P_0}(1)$ for all $P \in G$.
(ii) We have $h(2P) = 4h(P) + O(1)$ for all P.
(iii) For each $B \in \mathbb{R}$, the set $\{P \in G : h(P) \leq B\}$ is finite.

Proposition 42.1. *If G is any abelian group such that $G/2G$ is finite, and $h\colon G \to \mathbb{R}$ is any function satisfying (i) and (ii), then there exists $B > 0$ such that G is generated by $\{P \in G : h(P) \leq B\}$. So if h also satisfies (iii), then G is finitely generated.*

Proof. Let R be a set of coset representatives for $G/2G$. We will apply (i) only to $P_0 \in R$, and R is finite, so all the $O(1)$'s are uniformly bounded.

Given $Q_0 \in G$, we may write $Q_0 = 2Q_1 + r_1$ for some $Q_1 \in G$ and $r_1 \in R$; then

$$4h(Q_1) + O(1) = h(2Q_1) \leq 2h(Q_0) + O(1),$$

so

$$h(Q_1) \leq \frac{1}{2}h(Q_0) + O(1) \leq \frac{2}{3}h(Q_0);$$

if $h(Q_0)$ is sufficiently large. Choose B so that this holds whenever $h(Q_0) > B$. Let $S := \{P \in G : h(P) \leq B\}$. We may increase B if necessary to assume that $R \subseteq S$. Let $\langle S \rangle$ be the subgroup of G generated by S.

We claim that $\langle S \rangle = G$. Suppose that $Q_0 \in G$. If $h(Q_0) > B$, write $Q_0 = 2Q_1 + r_1$ as above. If $h(Q_1) > B$, repeat the process to write $Q_1 = 2Q_2 + r_2$, and so on. Since the height is shrinking by a constant factor each time, eventually we reach a Q_n with $h(Q_n) \leq B$, i.e., with $Q_n \in S$. (This is "Fermat's method of infinite descent"!) Now $Q_{n-1} = 2Q_n + r_n \in \langle S \rangle$, and $Q_{n-2} = 2Q_{n-1} + r_{n-1} \in \langle S \rangle$, and so on, until we show that $Q_0 \in \langle S \rangle$. This holds for every Q_0, so $\langle S \rangle = G$. □

The weak Mordell-Weil theorem combined with the fact that $h\colon E(\mathbb{Q}) \to \mathbb{R}$ satisfies the hypotheses of Proposition 42.1 proves that $E(\mathbb{Q})$ is finitely generated.

43. Faltings' theorem

The following was conjectured by Mordell in 1922, proved by Faltings in 1983, and reproved by a different method by Vojta in 1991.

Theorem 43.1. *Let X be a nice curve of genus $g > 1$ over $\mathbb{Q}$. Then $X(\mathbb{Q})$ is finite.*

Both proof methods are very difficult. With a lot of work, each can be used to get an upper bound on $\#X(\mathbb{Q})$, but neither gives a method to determine $X(\mathbb{Q})$ explicitly.

Acknowledgements

These notes are based partly on material in [Kob84], [Ser73], [Lan02], [Har77], and [Sil92].

References

[Har77] Robin Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977. Graduate Texts in Mathematics, No. 52. MR0463157 (57 #3116) ↑43

[Kob84] Neal Koblitz, *p-adic numbers, p-adic analysis, and zeta-functions*, 2nd ed., Graduate Texts in Mathematics, vol. 58, Springer-Verlag, New York, 1984. MR754003 (86c:11086) ↑43

[Lan02] Serge Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002. MR1878556 (2003e:00003) ↑43

[Ser73] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York, 1973. Translated from the French; Graduate Texts in Mathematics, No. 7. MR0344216 (49 #8956) ↑43

[Sil92] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, vol. 106, Springer-Verlag, New York, 1992. Corrected reprint of the 1986 original. MR95m:11054 ↑37.5, 43

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

Email address: poonen@math.mit.edu

URL: http://math.mit.edu/~poonen/