# Introduction to automorphic representations
## (DRAFT!)

Yiannis Sakellaridis

October 20, 2014

Disclaimer: These notes have been stitched together from lecture notes that I wrote for graduate classes that I taught at Rutgers-Newark and at Tel Aviv University, with relatively few subsequent edits. Since the beginning of October 2014 I have started editing them again, slowly, and the date of the latest update is the one you see on the first page. They have not been carefully checked for correctness, and they are very incomplete. Read at your own risk!

Meanwhile, by reading them, you agree to the following rules:

1. You will not print them out, except for small parts. Paper is too precious to waste on such an incomplete, constantly changing manuscript!

2. You will let me know of any mathematical mistakes that you encounter. I also welcome any other suggestions/ideas/questions. My email address is: sakellar@rutgers.edu.

# Contents

# Chapter 0

# Introduction

The purpose of these notes is to provide an introduction to the theory of automorphic representations and the Langlands program.

# Part I

# Basic notions: Algebraic geometry

This part has been written for people who come to the field with an analytic background, and should be skipped by all who are already familiar with basic concepts of algebraic geometry. Obviously, there are better places to get an introduction to algebraic geometry; I just wanted to provide a crash introduction to some basic notions that may be used elsewhere in these notes.

# Chapter 1

# The language of algebraic geometry: from rings to spaces

## 1.1 References

- Ravi Vakil's notes on Math 216 at Stanford: http://math.stanford.edu/~vakil/216blog/
- Hartshorne's "Algebraic Geometry"

## 1.2 Informal discussion

Rings show up everywhere in mathematics, for instance consider the following examples:

$$\mathbb{C}$$
$$\mathbb{C}[x]$$
$$\mathbb{C}[x,y]$$
$$\mathbb{C}[x,y]/(x^2-3y)$$
$$H(\mathbb{C}) \text{ (the ring of holomorphic functions on } \mathbb{C})$$
$$C(\mathbb{R}^2) \text{ (the ring of continuous functions on } \mathbb{R}^2)$$
$$\mathbb{Q}$$
$$\mathbb{Q}[x]$$
$$\mathbb{Z}$$
$$\mathbb{Z}[x,y]$$
$$\mathbb{C}[x]/x^2.$$

The basic philosophy of algebraic geometry is to consider every ring $R$ as *a space of functions on some space* $X$. But, what is the space, and how do we

distinguish between different types of functions? Today we will be a bit vague about the space, and will just focus on a few examples trying to understand the nature of this space and these functions intuitively.

For example, it is clear that in the case of $R = \mathbb{C}$, $R = \mathbb{C}[x]$, $R = \mathbb{C}[x, y]$ the space $X$ can be thought of as a point, a (complex) line and a (complex) two-plane, respectively. The functions here are polynomial (also called *regular*) functions on this space.

The spaces $H(\mathbb{C})$ and $C(\mathbb{R}^2)$ can also be thought of as functions on a complex line. Therefore, we see that the space itself does not suffice to describe the ring; we need to remember the ring. However, the space gives us a good picture about the structure of the ring, as we will see in the next paragraph.

Before we move to the next paragraph, an obvious question: OK, it was easy to describe a space for these rings. What about the rest? For instance, what about the rings $\mathbb{Q}, \mathbb{Z}, \mathbb{C}[x]/x^2$, etc., are they "functions" on some "space"? We will answer this in the next lecture.

The next paragraph, however, will give an answer about rings of the form: $\mathbb{C}[x, y]/(x^2 - 3y)$: this is just the space of regular functions on the subspace of $\mathbb{C}^2$ given by the equation: $x^2 - 3y = 1$.

## 1.3 Subspaces and (radical) ideals

The notion of a (closed) subspace also depends on the ring: the subspace has to be "cut out" by "functions" in the ring. For instance, if we are discussing the ring $\mathbb{C}[x, y]$ then the set of all $(x, y) \in \mathbb{C}^2$ satisfying the equation: $x^2 - 3y = 1$ is a valid subspace, but the subspace of all $(x, y)$ satisfying $x = e^y$ is not. However, the latter *is* a valid subspace when we discuss the ring $H(\mathbb{C}^2)$ (holomorphic functions). Again, the notion of "subspace" will be precisely defined later today, for some cases (see the paragraph on the Nullstellensatz), and in general in the next lecture. Let us discuss intuitively here about some basic properties:

Given a "subspace" $Y \subset X$, let $I(Y) \subset R$ denote the set of $f \in R$ which vanish on all of $Y$.

**Lemma 3.1.** *$I(Y)$ is an ideal.*

**Lemma 3.2.** *$I(Y)$ is a* radical *ideal, that is: if $f^n \in I(Y)$ for some $n$ then $f \in I(Y)$.*

**Lemma 3.3.** *If $Y_1 \subset Y_2$ then $I(Y_1) \supset I(Y_2)$.*

*Exercise.* Prove these lemmas rigorously for the cases where the space $X$ was defined above. Give examples of subspaces $Y$ and the corresponding radical ideals $I(Y)$.

*Exercise.* Define the notion of a *noetherian topological space*, so that it corresponds to the notion of a noetherian ring.

## 1.4 Morphisms

We continue our intuitive discussion with morphisms between spaces $X_1, X_2$ associated to two rings $R_1, R_2$. It is clear, again, that the notion of a morphism cannot be an arbitrary set-theoretic map: $X_1 \to X_2$ but one of the appropriate type, e.g. if $R_1, R_2$ are rings of polynomials then the morphism should be "given by polynomial equations". For example, $x \mapsto x^2$ is a valid morphism from $\mathbb{C} \to \mathbb{C}$ when $X_1 = X_2 = \mathbb{C}$, associated to the ring $R_1 = R_2 = \mathbb{C}[x]$, but $x \mapsto e^x$ is not a valid morphism for these rings.

Basic property of morphisms: A valid morphism $m : X_1 \to X_2$ gives rise to a *pull-back of functions* $m^* : R_2 \to R_1$, given by:

$$(m^* f)(x_1) = f(m(x_1)).$$

*Exercise.* Describe the morphism $R_2 \to R_1$ induced by the map $\mathbb{C} \ni x \mapsto x^2 \ni \mathbb{C}$, when $R_1 = R_2 = \mathbb{C}[x]$.

*Important remark:* How are we going to end up defining rigorously what a "valid morphism" is or, in the previous paragraphs, a "valid subspace" etc.? The answer is surprising: a valid morphism will be *identified* with the map between rings that, intuitively, it is supposed to induce. That is, *we think about spaces but we are using rings and algebra to define everything.*

*Example* 4.1. The map $m : x \mapsto e^x$ is clearly a valid map from $\mathbb{C} = X_1 \to X_2 = \mathbb{C}$ endowed with the rings $R_1 = R_2 = H(\mathbb{C})$, with associated pull-back:

$$(m^* f)(x) = f\left(e^x\right).$$

However, it is *also* a valid map when $X_1$ is endowed with the ring $R_1 = H(\mathbb{C})$ and $X_2$ is endowed with the ring $R_2 = \mathbb{C}[x]$. Why?

## 1.5 Non-reduced rings, non-radical ideals

Continuing with the same notation, observe the following: the whole space $X$ does not necessarily correspond to the "zero" ideal of $R$. In general, it corresponds to the ideal $\sqrt{(0)}$, where $\sqrt{\ }$ denotes the radical of an ideal. This lemma is obvious:

**Lemma 5.1.** *For a ring $R$ and an ideal $\mathfrak{a}$, the following are equivalent:*

- $\sqrt{\mathfrak{a}} = \mathfrak{a}$ *(i.e. $\mathfrak{a}$ is a radical ideal);*
- *$R/\mathfrak{a}$ has no nilpotent elements.*

A ring without nilpotent elements is called a *reduced ring*. The corresponding scheme (which we haven't yet defined) is called a *reduced scheme*.

If $\mathfrak{a}$ is a radical ideal in a polynomial ring, then we can think of the ring $R = k[x_1, \ldots, x_n]/\mathfrak{a}$ as the ring of regular (polynomial) functions on its "zero set" $Z(\mathfrak{a})$. If $\mathfrak{a}$ is not radical, how should we think of $R$?

*Example* 5.2. The ideal $(x^2-3y) \subset k[x,y]$ is radical. The ring $R = k[x,y]/(x^2-3y)$ is the ring of (regular) functions on the subspace $X$ of $k^2$ given by the equation $x^2 - 3y = 0$. Notice that any two elements of $k[x,y]$ are identified in $R$ if and only if they coincide on the points of $X$; in other words, by passing from $k[x,y]$ to $R$ we *remember* only the restriction of a function to $X$ and *forget* everything else about it.

*Example* 5.3. Now, let us consider the ring $R = k[x]/x^2$. It has a unique radical ideal, namely the ideal $(x)$; it is the radical of the zero ideal. Its zero set (more precisely, the zero set of its preimage in $k[x]$, considered as a space of functions on $k$) is the subset $X = \{0\}$ of $k$. Is it correct, thus, to think of $R$ as the space of functions on $X$? *No*, that would be the ring $k[x]/x(\simeq k)$. The ring $R$ remembers *more* from a function in $k[x]$; except for its restriction on $X$ it remembers *the first derivative* at that point. For example, the functions $3x+1$ and 1 are identified in $k[x]/x$, but *not* in $R$. On the other hand, $R$ does not remember the second derivative, e.g. the functions $3x+1$ and $x^2+3x+1$ are identified in $R$.

The way that we usually think of the "space" underlying $X$ is a "fattened" version of $X$, namely $X$ together with an "infinitesimal neighborhood" of "first order". If we were considering, instead, the ring $k[x]/x^3$, it would be an infinitesimal neighborhood of "second order" (in the sense that it remembers the second derivative), and so on.

*Exercise.* How would you describe the underlying space of $R = k[x,y,z]/(x-y,(y-z^2)^2)$?

## 1.6 Nullstellensatz

Now we start the rigorous discussion, but to do this we will restrict ourselves to a very small class of all the examples of rings that we mentioned: we will talk only about *finitely generated $k$-algebras*, where $k$ is some *algebraically closed* field (think of $\mathbb{C}$, if you prefer, at first reading). Hence, such a ring $R$ is the quotient of $k[x_1,\dots,x_n]$, for some $n$, by an ideal.

In fact, let us start with $R = k[x_1,\dots,x_n]$. Before continuing, return to the previous paragraphs and check all the statements for this ring (making them precise whenever they are not).

For any subset $S \subset R$, we let $Z(S)$ denote the *zero set* of $S$, that is the set of points $x \in k^n$ such that $f(x) = 0$ for all $f \in \mathfrak{S}$. Check the following:

**Lemma 6.1.** *If $\mathfrak{a}$ is the ideal generated by $S$ then $Z(S) = Z(\mathfrak{a})$.*

**Definition.** The subsets of $k^n$ of the form $Z(S)$ are called *algebraic sets.*

As we will see, they form the closed subsets for a topology (the Zariski topology) on $k^n$, so in the future we will not be using the term "algebraic set" but instead we will be saying "Zariski closed subset".

We are ready to formulate the Nullstellensatz (deep theorem, will not prove here, consult an algebra book like Lang's) :

**Theorem 6.2** (Hilbert's Nullstellensatz)**.** *If $f \in R$ vanishes on $Z(\mathfrak{a})$ (i.e. $f(x) = 0$ for all $x \in Z(\mathfrak{a})$ then there is an integer $r$ such that $f^r \in \mathfrak{a}$.*

*Hence, the maps: $\mathfrak{a} \mapsto Z(\mathfrak{a})$ and $Y \mapsto I(Y)$ give rise to a bijection between* radical ideals *and* algebraic subsets *of $k^n$.*

In particular, *maximal ideals* are in bijection with *points in $k^n$*. Explicitly, the maximal ideal $\mathfrak{m}_a$ corresponding to the point $a = (a_1, a_2, \ldots, a_n)$ is the ideal $(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$, and the homomorphism:

$$R \to R/\mathfrak{m}_a = k$$

is "evaluation at $a$".

*Exercise.* Show that the analogous theorem fails for the ring $C(k^n)$ of continuous functions!

Now let us consider the more general situation of a finitely generated $k$-algebra, i.e. $R = k[x_1, \ldots, x_n]/\mathfrak{a}$, where $\mathfrak{a}$ is some ideal. First of all, observe:

**Lemma 6.3.** *Ideals of $R$ are in natural bijection with ideals of $k[x_1, \ldots, x_n]$ containing $\mathfrak{a}$.*

It is immediate then to apply the Nullstellensatz in order to generalize it to $R$. Notice that we should be thinking of $X := Z(\mathfrak{a})$ as the "underlying space" of $R$:

**Theorem 6.4.** *There is a natural bijection between radical ideals of $R$ and algebraic subsets of $X$.*

*Remark.* It is not nice that the description of $X$ depends on the presentation of $R$, i.e. the way we choose to realize the abstract ring $R$ as a quotient of a polynomial ring. This will be fixed when we talk about schemes, but we note here that the solution will go through the observation that we made in **??**, and which generalizes here:

> Points of $X$ are in bijection with maximal ideals of $R$.

## 1.7 The maximal spectrum, and the Zariski topology

We have seen that points on $k^n$ are in bijection with maximal ideals in $R$. What follows holds for every ring $R$, not necessarily those of the previous paragraph. To give a completely analogous definition, denote by $\operatorname{spec}_M R$ the set of maximal ideals of $R$ (the *maximal spectrum* of $R$), and given any $S \subset R$ denote by $Z(S) \subset \operatorname{spec}_M R$ the set of zeroes of $S$, i.e. the set of maximal ideals which contain $S$.

**Lemma 7.1.** *1. Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals in $R$, then $Z(\mathfrak{ab}) = Z(\mathfrak{a}) \cup Z(\mathfrak{b})$.*

2. *Let* $(\mathfrak{a}_i)_{i\in I}$ *be any (possibly infinite) collection of ideals, then* $Z(\cup_{i\in I}\mathfrak{a}_i) = \cap_{i\in I} Z(\mathfrak{a}_i)$.

*Proof.* 1. We have $\mathfrak{ab} \subset \mathfrak{a} \cap \mathfrak{b}$, so the direction $\supset$ is obvious. Vice versa, if $\mathfrak{m}$ is a maximal (in fact, prime) ideal which contains $\mathfrak{ab}$ then it must contain $\mathfrak{a}$ or $\mathfrak{b}$. (Otherwise, there are elements $f_1, f_2 \in \mathfrak{a}, \mathfrak{b}$, respectively, which don't belong to $\mathfrak{m}$, but $f_1 f_2 \in \mathfrak{ab} \subset \mathfrak{m}$, contradiction since $\mathfrak{m}$ is prime.) Therefore, $\mathfrak{m} \in Z(\mathfrak{a}) \cup Z(\mathfrak{b})$.

2. Clear.

□

**Corollary 7.2.** *The subsets of the form* $Z(S)$, $S \subset R$, *form the closed sets for a topology on* $\operatorname{spec}_M R$ *(called the* Zariski topology*).*

## 1.8 Irreducibility

In a topological space $X$, a subset $A$ is called *irreducible* if it cannot be written as the disjoint union of two proper, closed subsets. This notion is not very useful in the usual, Hausdorff topologies, but there is a lot of irreducible sets in the Zariski topology:

**Lemma 8.1.** *Let* $\mathfrak{a}$ *be a radical ideal in a ring* $R$. *The set* $A = Z(\mathfrak{a})$ *is irreducible if and only if* $\mathfrak{a}$ *is prime.*

*Proof.* If $A = B \cup C$ with $B = Z(\mathfrak{b})$ and $C = Z(\mathfrak{c})$ proper, closed subsets, then $\mathfrak{b} \cdot \mathfrak{c} \subset \mathfrak{a}$, but $\mathfrak{b} \supsetneq \mathfrak{a}$ and $\mathfrak{c} \supsetneq \mathfrak{a}$. As we saw in the proof of Lemma **??**, we can find such ideals if and only if $\mathfrak{a}$ is not prime. (The "if" part comes from the definition of a prime ideal; if $\mathfrak{a}$ is not prime then there are elements $f_1, f_2 \notin \mathfrak{a}$ such that $f_1, f_2 \in \mathfrak{a}$, and we can take the ideals generated by $\mathfrak{a}$ and each of the $f_i$'s.) □

# Chapter 2

# Affine schemes

## 2.1 The category of affine schemes

**Definition.** The category of *affine schemes* is the opposite category of the category of *rings*.

In other words, any affine scheme corresponds to a ring and vice versa, and a morphism of affine schemes is the same as a morphism of rings, except that we write arrows in the opposite direction. Whatever we say about affine schemes corresponds to a statement about rings, there is absolutely no difference. It's just that by inverting the arrows we will get a more geometric picture of rings, which we can later generalize to (non-affine) schemes.

## 2.2 The underlying space of an affine scheme

In the previous lecture we never defined rigorously the underlying topological space of a ring, except in the cases where the Nullstellensatz applies, in which case we saw a certain space (whose points were in bijection with *maximal* ideals of the ring), and endowed it with a topology (the Zariski topology).

What we will do here is not completely analogous to this: following[1] Grothendieck, we will define a space whose points will correspond not only to maximal ideals of the ring, but to *all prime ideals*. Let us recall what prime ideals correspond to: they correspond to *irreducible, closed subspaces* of our space, maybe with fattened neighborhoods if the ideal is not reduced.

But here is a problem: if our space is to consist of both the maximal and the non-maximal ideals, how will the inclusion relations be expressed? For example, since the point $(1, 2)$ of $\mathbb{C}^2$ belongs to the subspace $2x - y = 0$, but both the

[1] The Wikipedia article on schemes has interesting information on the history of this idea: It was Krull in the 1930s who defined the topological space consisting of all primes of a ring, but he abandoned the idea since it didn't seem interesting to his peers. By the time that Grothendieck gave the full definition of a scheme, the underlying topological space had been used by Serre, Chevalley and Nagata.

point $(1,2)$ and the subspace $2x - y = 0$ will be "points" of our new space (since they correspond to distinct prime ideals), how will the inclusion relation be expressed?

The answer is that the topology will be a weird kind of topology, where not all points are closed! For example, the "point" corresponding to the ideal $(2x-y)$ will not be closed, and will contain in its closure the "point" corresponding to the ideal $(1,2)$. Actually, the *closed* points will precisely be those corresponding to *maximal* ideals, and this makes sense: Since they don't belong to any larger ideal, they shouldn't contain any subspaces in their closure; they correspond to our conventional notion of point.

**Definition.** Given an affine scheme $X$, the set of *points* of $X$ is the set of prime ideals of the corresponding ring. We endow this set with a topology (Zariski topology), according to which a point $\mathfrak{p}$ belongs to the closure of a set $S$ if and only if the ideal $\mathfrak{p}$ contains all ideals contained in $S$.

By abuse of notation, we also denote the set of points of $X$ by $X$. Here is the important, though obvious, fact:

**Lemma 2.1.** *A morphism of schemes $X_1 \to X_2$ also induces a set-theoretic map: $X_1 \to X_2$.*

This *would not be true* if we had just included maximal ideals in the set. For example, for the natural morphism: $\operatorname{spec} k(X) \to \operatorname{spec} k[X]$ the zero ideal in $k(X)$, which is maximal because $k(X)$ is a field, goes to the zero ideal in $k[X]$, which is not maximal.

The proof that this is indeed a topology is the same as with $\operatorname{spec}_M$ that we saw in the previous lecture. (Notice that we never used the fact that those ideals were maximal; just prime.) Observe:

- Points are not necessarily closed (it is not a $T_1$ *topological space*); the closure of a point-prime $\mathfrak{p}$ is the set of all primes containing it.
- Thus, *closed points* are precisely the *maximal ideals*.

*Example* 2.2. The space $\operatorname{spec} k[X,Y]$ has a *generic point* $\xi := (0)$, i.e. a point whose closure is the whole space; closed points $\mathfrak{m}_{a,b} = (x-a, x-b)$; and those primes $\mathfrak{p} = (P)$, where $P$ is an irreducible polynomial. It turns out that these are all its points, for reasons that we will discuss later. The closure of $\mathfrak{p} = (P)$ contains itself and the points $\mathfrak{m}_{a,b}$ with $P(a,b) = 0$.

*Example* 2.3. $\operatorname{spec} \mathbb{Z}$ has a generic point, and one closed point for each prime number.

As in the discussion of algebraic subsets of $k^n$ (when $k$ was an algebraically closed field), the zero set of an ideal is equal to the zero set of its radical. More is true, namely that as in the Nullstellensatz the zero set determines the radical ideal (but now *we have to include all primes in the zero set for that to be true*):

**Lemma 2.4.** *Let $A$ be any ring. An element of $A$ is nilpotent if and only if it lies in every prime ideal. For any ideal $\mathfrak{a}$ we have:*

$$\sqrt{\mathfrak{a}} = \cap_{\mathfrak{p} \supset \mathfrak{a}} \mathfrak{p},$$

*where $\mathfrak{p}$ stands for a prime ideal.*

*Proof of the lemma.* If $f^n = 0$ then $f^n \in \mathfrak{p}$ for every prime $\mathfrak{p}$ and hence $f \in \mathfrak{p}$ since $\mathfrak{p}$ is prime. Vice versa, if $f^n \neq 0$ then $S := \{1, f, f^2, f^3, \dots\}$ is a multiplicative set not containing zero, and then it is known that there is a prime ideal which doesn't intersect $S$ (exercise!). The second assertion follows from applying the first to the ring $A/\mathfrak{a}$. □

Finally, with exactly the same proof as before we have:

**Lemma 2.5.** *For a radical ideal $\mathfrak{a}$, the zero set $Z(\mathfrak{a})$ is* irreducible *if and only if $\mathfrak{a}$ is prime.*

Reminder: irreducible means that it's not the union of two proper closed subsets.

## 2.3 Localization

The basic reason why thinking of rings geometrically is so useful is that many properties of rings are completely local in nature, in the sense that they are preserved if we restrict to, say, a Zariski open subset. We explain here this notion of "restriction".

Given a multiplicative (i.e. closed under multiplication) subset $S$ of a ring $R$ we define the *localization* of $R$ at $S$, denoted $R[S^{-1}]$, to be the initial object among homomorphisms of rings $R \to M$ which take $S$ to units. In other words, it is the universal ring, together a canonical map $R \to R[S^{-1}]$ which takes elements of $S$ to units, with the property that any other homomorphism $R \to M$ taking $S$ to units factors through a map: $R[S^{-1}] \to M$.

It is easy to see that the localization can be constructed as follows: Its elements consist of pairs $(r, s), r \in R, s \in S$, modulo the following equivalence: $(r_1, s_1) \sim (r_2, s_2)$ iff there exists an $s \in S$ with $s(r_1 s_2 - r_2 s_1) = 0$. We will denote the pair $(r, s)$ by $r/s$, but the condition of equivalence is a bit stronger than the usual equality of fractions because, remember, we want elements of $S$ to become units, and if $s$ times something is zero then that something should be zero in our new ring.

**Lemma 3.1.** *The kernel of $R \to R[S^{-1}]$ is the set of all elements of $R$ which are annihilated under multiplication by an element of $S$.*

In particular, for division rings the map is injective.

*Proof.* Exercise! □

*Example* 3.2. Let $R$ be any ring and $f$ an element which is not nilpotent. Let $S = \{f, f^2, f^3, \dots, \}$, then $R[S^{-1}] = R[f^{-1}] := R[X]/(Xf - 1)$.

Geometrically, inverting some elements means removing their zero set from our scheme, since their zero set consists of all ideals that contain them, and since our element becomes invertible these ideals generate the whole ring in $R[S^{-1}]$. It is in this sense that localization corresponds to "restriction to a Zariski open set" (but also to much more, as we will see). The following proposition explains that:

**Proposition 3.3.** *Primes of $R[S^{-1}]$ are in bijection with primes of $R$ not meeting $S$. (Each prime in $R[S^{-1}]$ being the prime generated by its intersection with $R$.)*

*Proof.* Let $\phi$ denote the morphism $R \to R[S^{-1}]$. If $\mathfrak{p} \in \operatorname{spec} R[S^{-1}]$ then $\phi^{-1}(\mathfrak{p})$ does not meet $S$ because $\mathfrak{p}$ doesn't meet $\phi(S)$ (all elements of $\phi(S)$ are units).

If $P \in \operatorname{spec} R$ doesn't meet $S$, we claim that $\phi(P)$ generates a prime ideal. Indeed, if $(r_1 r_2, s_1 s_2)$ belongs to the ideal generated by $\phi(P)$ this means that there exist $p \in P$ and $s \in S$ such that:

$$(p, s) \sim (r_1 r_2, s_1 s_2),$$

i.e. there is an $s' \in S$ such that: $s' p s_1 s_2 = s' r_1 r_2$. The element on the left is in $P$, and since $s'$ is not in $P$ it means that $r_1 \in P$ or $r_2 \in P$, which implies the claim.

Finally, the same argument shows that if $(r, s)$ belongs to the ideal $\phi(P)[S^{-1}]$ generated by $\phi[P]$ then $r \in P$, so $P = \phi^{-1}(\phi(P)[S^{-1}])$. □

*Example* 3.4. Let $R = \mathbb{C}[x, y]/(xy)$, and let $S = x, x^2, x^3, \dots$. Then $R[S^{-1}] = \mathbb{C}[x, x^{-1}]$. What is the geometric picture?

*Example* 3.5. If $R$ is an integral domain and $S = R \setminus \{0\}$ then $K(R) := R[S^{-1}]$ is the quotient field. Notice that in terms of schemes the map $\operatorname{spec} K(R) \to \operatorname{spec} R$ is simply the inclusion of the generic point.

*Example* 3.6. If $R$ is an arbitrary ring and $S =$ the set of non-zero divisors, then $K(R) := R[S^{-1}]$ is called the *total quotient ring* of $R$.

## 2.4 Localization at a prime

Here is the most basic use of localization: let $\mathfrak{p}$ be a prime of $R$, then $S = R \setminus \mathfrak{p}$ is a multiplicative set. The localization $R_{\mathfrak{p}} := R[S^{-1}]$ is called the *localization of $R$ at $\mathfrak{p}$*.

*Example* 4.1. The localization of $\mathbb{Z}$ at $(p)$ consists of all rational number which do not have a power of $p$ in their denominator.

*Example* 4.2. The localization of $\mathbb{C}[X]$ at $(0)$ consists of all rational functions which do not have a power of $X$ in the denominator.

$R_{\mathfrak{p}}$ is a local ring, i.e. it has a unique maximal ideal, and this maximal ideal "is $\mathfrak{p}$" (i.e. it is the ideal generated by $\mathfrak{p}$, but we will still denote it by $\mathfrak{p}$). More precisely, it follows from Proposition **??**:

**Lemma 4.3.** *Primes of $R_{\mathfrak{p}}$ are in natural bijection with primes of $R$ which are contained in $\mathfrak{p}$. (Geometrically, points of* spec $R$ *which contain $\mathfrak{p}$ in their closure.)*

This localization is of a slightly different nature than the example of inverting an element $f \in R$: in that example, we localized away from the zero set of $f$, i.e. we restricted to an open subset of our Zariski space. Now, given $\mathfrak{p}$ (whose closure is an irreducible subset $Y$ of our space), we feel free to remove *any closed subset that doesn't contain $Y$*. This means that our new functions are allowed to be defined only in some open set, *as long as $Y$ is not in the complement of this open set*. Those functions are even allowed to have poles on $Y$, as long as they don't on all $Y$ (indeed, all prime ideals larger than $Y$, i.e. all proper irreducible subspaces of $Y$, have been erased).

*Example* 4.4. Localizing $\mathbb{C}[x, y]$ at $(x)$ we get all rational functions of $x$ and $y$ whose denominator is not divisible by $x$. This admits a homomorphism to $\mathbb{C}(y)$, namely setting $x = 0$. Geometrically, this homomorphism is restriction of our functions to the $y$-axis.

# Chapter 3

# Sheaves and schemes

Later.

# Chapter 4

# Noetherian rings

## 4.1 References

- S. Lang, "Algebra", Chapter X.
- D. Eisenbud, "Commutative Algebra with a view towards Algebraic Geometry", Chapter 3.
- Ravi Vakil's notes for 216, chapters 6 and 12.

## 4.2 Recollection of definitions and basic properties

A module for a ring $R$ is *noetherian* if every submodule is finitely generated, and a ring is *noetherian* if it is Noetherian as a module over itself, i.e. every ideal is finitely generated. Equivalently, if any increasing sequence of submodules (ideals in the case of the ring) stabilizes. A finitely generated module for a noetherian ring is noetherian.

*Hilbert's Basis Theorem* states that if $R$ is noetherian then so is $R[x]$. Obviously, quotients of noetherian modules are noetherian, hence every finitely generated ring over a noetherian ring is noetherian.

## 4.3 Primary decomposition and associated ideals

An ideal $I$ is called *primary* if $ab \in I$ implies $a \in I$ or $b^n \in I$ for some integer $n$. Equivalently, if and only if all zero divisors of $R/I$ are nilpotent.

**Lemma 3.1.** *The radical of a primary ideal is prime. We say that* $I$ belongs to *the prime* $\sqrt{I}$.

*Proof.* Immediate from the definitions. □

*Remark.* Let $\mathfrak{p} = \sqrt{I}$, $I$ primary. If $ab \in I$ then there are two possibilities:

- either one of them, say $a$ is not in $\mathfrak{p}$, in which case $b \in I$;
- or $a^n \in I$ *and* $b^m \in I$ for some $m, n$.

Indeed, if $a \notin \mathfrak{p}$ then no power of $a$ can be in $I$ and hence $b \in I$. On the other hand, if $a \in \mathfrak{p}$ then some power of $a$ is in $I$, because $\mathfrak{p}$ is the radical of $I$.

Using this fact, you can easily prove:

**Lemma 3.2.** *The intersection of a finite number of primary ideals belonging to the same prime is primary.*

*Example* 3.3. The primary ideals of $\mathbb{Z}$ are $(0)$ and $(p^n)$, where $p$ ranges over all prime numbers.

*Example* 3.4. The ideal $(x, y^2) \subset \mathbb{C}[x, y]$ is primary.

A geometric way to think about a radical ideal is that the "underlying space" (the topological space $\operatorname{spec}(R/I)$) is irreducible, but it may be "fat" (not reduced). Well, not quite, actually: we said that the radical of a primary ideal is prime, but that doesn't mean that every ideal whose radical is prime is primary. The example below has an irreducible underlying space (i.e. a prime radical), but is not primary:

*Example* 3.5. The ideal $I = (x^2, xy)$ is not primary, because $xy$ belongs to it, but $x$ doesn't, neither does $y^n$. The radical if $I$ is $(x)$, which is prime.

What is happening here? The "fattening" along the line $x = 0$ is not the same as the "fattening" at the point $(0, 0)$. Thus, even if the underlying topological space is irreducible, we can still distinguish between "embedded components", one contained in the other, with distinct fattenings. To make this rigorous in an example, show that:

$$(x^2, xy) = (x) \cap (x^2, xy, y^n), \tag{4.1}$$

and that these two ideals are primary with radicals: $(x)$ and $(x, y)$, respectively.

The big theorem (actually, not too hard, but will not prove) for noetherian rings is:

**Theorem 3.6.** *Every ideal is the intersection of a finite number of primary ideals. If:*

$$I = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_k$$

*is a* minimal *decomposition, i.e. one where the primes* $\mathfrak{p}_i = \sqrt{\mathfrak{q}_i}$ *are distinct and there are no redundancies, then:*

- *the set of primes* $\mathfrak{p}_i$ *(the* associated primes *of* $I$*) is uniquely determined by* $I$*;*

- *the set of primary ideals* $\mathfrak{q}_i$ *belonging to* minimal *associated primes is uniquely determined by* $I$.

As an example, the decomposition (**??**) is such a decomposition, but it is not unique: we could also have written:

$$(x^2, xy) = (x^2, x(x+y), (x+y)^n).$$

The non-minimal associated primes are called *embedded primes*. They correspond to these "embedded fattenings" that we discussed above.

Further facts:

- The associated primes of $I$ are precisely those primes which are annihilators of elements of $R$ modulo $I$.
- The minimal primes associated to $I$ are precisely the minimal primes containing $I$.

By *associated primes of* $R$ we mean the associated primes of the ideal (0). These correspond to the irreducible components, as well as the "embedded components" of spec $R$ that were discussed above.

If associated primes of $R$ are annihilators of some elements of $R$, what about annihilators of arbitrary elements? Write $\mathrm{ann}(f)$ for the set of elements annihilating $f$. (If $f \neq 0$ then this is a proper ideal, and it is the zero ideal if and only if $f$ is not a zero divisor.)

**Lemma 3.7.** *Given* $f \neq 0$ *there is an associated prime* $\mathfrak{p}$ *of* $R$ *such that* $\mathrm{ann}(f) \subset \mathfrak{p}$.

Intuition: if $f \neq 0$ there must be some "embedded component" (in the above sense) where the localization of $f$ is non-zero. (Recall: the element $f$ is in the kernel of $R \to R_{\mathfrak{p}}$ iff $\mathrm{ann} f \not\subset \mathfrak{p}$.) Notice that there needn't be an *irreducible component* where this localization is nonzero. For example, if $R = \mathbb{C}[x,y]/(x^2, xy)$ and $f = x$ then $\mathrm{ann}(x) = (x, y)$, in particular $\mathrm{ann}(x) \not\subset (x)$.

*Proof.* It is enough to show that an ideal which is maximal with the property of annihilating a non-zero element of $R$ is prime. Let $I = \mathrm{ann} f$ be such an ideal and $ab \in I$ with $a \notin I$. Then $abf = 0$ but $af \neq 0$, hence $(b, I) \subset \mathrm{ann}(af)$. Since $I$ was maximal, this implies that $b \in I$. □

**Corollary 3.8.** *1. The set of zero divisors of* $R$ *is precisely the union of associated primes.*

*2. The natural map* $R \to \prod R_{\mathfrak{p}}$ *is an injection, where the product is over all associated primes of* $R$.

*Proof.* Exercise! □

For the unproven statements in this section, cf. Eisenbud Theorem 3.10 and Corollary 3.5.

## 4.4 Dimension

The *height* of a prime ideal $\mathfrak{p}$ is the supremum of all integers $n$ such that there exists a sequence of distinct prime ideals: $\mathfrak{p}_0 \subset \mathfrak{p}_1 \subset \cdots \subset \mathfrak{p}_n = \mathfrak{p}$.

The *(Krull) dimension* of a ring $R$ is the supremum of heights of prime ideals in $\mathfrak{p}$. To say it geometrically: the supremum of numbers $n$ such that there exists a sequence of distinct irreducible algebraic subsets:

$$X_0 \supset X_1 \supset \cdots \supset X_n$$

in $\operatorname{spec} R$.

For *noetherian* rings this is a very reasonable notion of dimension, because it turns out to be the unique one satisfying the four axioms:

D1: (Dimension is a local property) $\dim R = \sup_{\mathfrak{p}} \dim R_{\mathfrak{p}}$.

D2: (Nilpotents do not affect dimension) $\dim R = \dim R^{\mathrm{red}}$.

D3: (Dimension is preserved by finite morphisms) For a finite morphism $\operatorname{spec} S \to \operatorname{spec} R$ we have $\dim S = \dim R$.

D4: If $k$ is a field, then $\dim k[[x_1, \ldots, x_r]] = r$, and if $R$ is a discrete valuation domain then $\dim R[[x_1, \ldots, x_r]] = r + 1$.

Given a ring $R$ and an irreducible algebraic subset $Y$ in $\operatorname{spec} R$ (i.e. a prime $\mathfrak{p}$ so that $Y = \{\widehat{\mathfrak{p}}\}$), the *dimension* of $Y$ is the dimension of $R/\mathfrak{p}$ and the *codimension* of $Y$ is the height of $\mathfrak{p}$; equivalently, the dimension of $R_{\mathfrak{p}}$.

Clearly,

$$\dim Y + \operatorname{codim} Y \leqslant \dim R.$$

However, as the following example shows we do not always have equality:

*Example* 4.1. Let $R = k[X_1, X_2, X_3]/(X_1)(X_2, X_3)$ (this represents the union of a plane and an intersecting line in 3-space). Let $\mathfrak{p}$ be the prime $(X_1 + 1, X_2, X_3)$; this corresponds to a point on the line, not contained in the plane. Its dimension is zero, its codimension one, but the dimension of $R$ is two.

However, we will see in the next paragraph a situation where they behave well.

We finish this discussion with a mention of *Krull's Hauptidealsatz* (Principal Ideal Theorem), which we will not prove. It provides the expected answer to the question: what is codimension of the zero set of an element of $R$?

**Theorem 4.2.** *Let $R$ be a noetherian ring, and let $f \in R$ be neither a zero divisor nor a unit. Then every irreducible component of the zero set $Z(f)$ (i.e. every* minimal associated prime of $(f)$*) has codimension one (has height one).*

The way this is usually formulated is that every prime minimal among those which contain $f$ has height 1. Notice that it is *not* true that $(f)$ cannot have *embedded* primes (necessarily of larger codimension), although it *is* true for

normal rings, as we will discuss later. A counterexample is the prime $\mathfrak{p} = (x, y, u, v)$ in the ring $R = \mathbb{C}[x, y, u, v]/(x, y) \cap (u, v)$. This is the union of two planes in $\mathbb{A}^4$ meeting at a point. You can check that $f = x + u$ is a non-zero divisor, and hence cuts out a subscheme of codimension one, but the prime $\mathfrak{p}$ is an associated prime for $(f)$ because it is the annihilator of the nonzero element $x - u$ modulo $(f)$.

A generalization of this theorem is the following:

**Theorem 4.3.** *Let $R$ be a noetherian local ring with maximal ideal $\mathfrak{m}$, then $\dim R$ is the minimal number $n$ such that there exist $n$ elements $f_1, \ldots, f_n \in \mathfrak{m}$ such that not all of them are contained in a prime other than $\mathfrak{m}$.*

# Chapter 5

# Noetherian rings of dimension one

*Every ring in this lecture is noetherian, and every scheme is covered by a finite set of open affine noetherian subschemes.*

We study schemes of dimension one not only for their intrinsic interest (they include Riemann surfaces, rings of integers in number fields etc.), but also because by the process of localization we often reduce problems about arbitrary schemes to schemes of dimension one (by localizing at minimal primes).

In particular, we can start understanding in this setting several notions of *smoothness*, and prepare ourselves for a general discussion of smoothness in a later lecture. The word "smoothness" is used colloquially here, as there is also a rigorous notion of "smoothness of a scheme *over* a base scheme".

## 5.1 UFDs and PIDs

**Proposition 1.1.** *For a ring of dimension one, the following are equivalent:*

1. *It is a unique factorization domain.*
2. *It is a principal ideal domain.*

*Proof.* The second implies the first for every ring.

Vice versa, assume that $R$ is a UFD, let $\mathfrak{p}$ be a non-zero prime and take $0 \neq f \in \mathfrak{p}$. Then $\mathfrak{p} \supset (f) = \prod_i (f_i)$ is a factorization into principal prime ideals, hence there is an $i$ such that the prime $(f_i)$ is contained in $\mathfrak{p}$. But the ring has dimension one, hence $\mathfrak{p} = (f_i)$.

Thus, all primes are principal. Now show that the non-principal ideals, if they exist, have maximal elements (exercise!). Since $I$ cannot be prime, there are $a, b \notin I$ with $ab \in I$. Then $(I, a)$ and $(I : a)$ are principal. ($(I : a)$ denotes the set of $x$ with $xa \in I$, and hence contains $b$, so is strictly larger than $I$.) Their product obviously is contained in $I$. On the other hand, if $i \in I \subset (I, a) = (c)$,

so $i = uc$, then $u \in (I : a) = (d)$ and hence $i \in (I, a)(I : a) = (cd)$. Hence $I = (cd)$, a contradiction. □

## 5.2 Normal and regular domains

Recall that a unique factorization domain is always integrally closed (i.e. contains all elements of the fraction field which are integral over it); an integrally closed domain is called *normal* in algebraic geometry jargon. On the other hand, a principal ideal domain has the obvious property that for every non-zero prime $\mathfrak{p}$, the $R/\mathfrak{p}$ space $\mathfrak{p}/\mathfrak{p}^2$ is one-dimensional (generated by the image of a generator of $\mathfrak{p}$). We will see that the space $\mathfrak{p}/\mathfrak{p}^2$ is the *cotangent space* at $\mathfrak{p}$, and the equality of its dimension with the dimension of the ring is the best intrinsic notion of smoothness, called *regularity*.

As a counterexample, consider the cuspidal curve $y^2 = x^3$. Show (exercise!) that the cotangent space at the point $(0, 0)$ is two-dimensional.

An analog of the above proposition for these more general notions are:

**Theorem 2.1.** *For a noetherian domain of dimension one, the following are equivalent:*

1. *It is normal.*
2. *It is regular.*

*Proof.* In the next subsection we will prove that the theorem is true for local rings.

Now, assume that $R$ is normal, and notice that $\mathfrak{p}/\mathfrak{p}^2 = \mathfrak{p}_\mathfrak{p}/\mathfrak{p}_\mathfrak{p}^2$ as $R/\mathfrak{p} = R_\mathfrak{p}/\mathfrak{p}_\mathfrak{p}$-vector spaces. Normality is preserved by localization, thus regularity follows from the corresponding local statement.

Vice versa, assume that $R$ is regular, then $R_\mathfrak{p}$ is regular for every prime $\mathfrak{p}$. We claim that $R = \cap_\mathfrak{p} R_\mathfrak{p}$; this will prove normality, because the intersection of normal domains is normal. Indeed, if $\frac{a}{b} \in K(R)$ with $a \notin (b)$, then applying Corollary **??** to the ring $R/(b)$ we deduce that there is an associated prime $\mathfrak{p}$ of $(b)$ such that $a \notin \mathfrak{p}$. But this means that $\frac{a}{b} \notin R_\mathfrak{p}$.

□

In the process of proving that normal implies regular for a local ring of dimension one, we will need the following result which does not use the local property and is important for its own sake:

**Proposition 2.2.** *Let $R$ be a noetherian normal domain of dimension one. Then every ideal is invertible, i.e. the $R$-submodule of $K(R)$ (=*fractional ideal*) $I^{-1} = \{x \in K(R) | xI \subset R\}$ satisfies:*

$$I \cdot I^{-1} = R.$$

Notice that by $I \cdot I^{-1}$ we denote *sums* of elements of the form $j \cdot i, j \in I^{-1}, i \in I$, so that the product of fractional ideals is again a fractional ideal.

(More generally, I guess, we would define in such a way the product of $I$ with any module, so that it is a submodule.)

*Proof.* First, we claim that a maximal non-invertible ideal is prime, thus it suffices to prove the proposition for prime ideals. Indeed, if $I$ is maximal non-invertible and $ab \in I$ with $J := (a, I) \supsetneq I$ then $J$ is invertible and hence there are elements $x, y \in K(R)$ and $i \in I$ such that $xa + yi = 1 \Rightarrow xab + yib = b$. Thus,

$$I \subset (I, b) \subset IJ^{-1} \subsetneq R,$$

thus, unless $(I, b) = I$, we get that $K = IJ^{-1}$ is invertible, $IJ^{-1}K^{-1} = R$, thus $I$ would be invertible with $I^{-1} \supset J^{-1}K^{-1}$, a contradiction. Therefore, $b \in I$ and hence $I$ is prime.

If $I$ is prime, we claim that at least there is an element $x \in K(R) \setminus R$ such that $xI \subset R$, i.e. $I^{-1} \supsetneq R$. Again, same flavor of argument: let $J$ be maximal among ideals in $I$ for which there is such an element, then we claim that $J$ is prime and hence $J = I$ by dimension one. Indeed, if $ab \in J$, $c \in J^{-1} \setminus R$ then $cb \cdot (a, J) \subset R$, thus if $a \notin J$ then by maximality $cb \in R$. But then $c \in (b, I)^{-1}$, and by maximality $(b, I) = I$.

Now let $I \neq 0$ and prime (for $I = 0$ the proposition is obvious), then $I^{-1}I$ is either $I$ or $R$ (since $I$ is maximal). We have arrived at the most important part of the proof: *if $xI \subset I$ then $x$ is integral over $R$.* By normality, this will imply that $x \in R$, a contradiction; thus $I^{-1}I$ has to be equal to $R$. The proof of this fact is contained in the lemma that follows (with $A = R$, $\alpha = x$, $M = I$). □

**Lemma 2.3.** *Let $A \subset B$ be rings and $\alpha \in B$. Then $\alpha$ is integral over $A$ if and only if there is a faithful $A[\alpha]$-module (i.e. a module $M$ such that the map $A[\alpha] \to \mathrm{End}(M)$ is injective) which is finitely generated as an $A$-module.*

*Proof.* If $\alpha$ is integral then $A[\alpha]$ is such a module.

Vice versa, if $w_1, \ldots, w_n$ are generators of $M$ over $A$ then there is an $n \times n$ matrix $K = (k_{ij})$ with coefficients in $A$ such that $\alpha w_i = \sum_{j=1}^{n} k_{ij} w_j$. Let $f \in A[X]$ be the characteristic polynomial of $K$. Then $f(\alpha) \in A[\alpha]$ satisfies $f(\alpha)M = 0$, and since $M$ is faithful this implies that $f(\alpha) = 0$. Hence $\alpha$ is the root of a monic polynomial. □

## 5.3 Local domains

**Theorem 3.1.** *For a noetherian local domain of dimension one, the following are equivalent:*

1. *It is normal.*
2. *It is a UFD.*
3. *It is a PID.*
4. *It is regular.*

*Proof.* We already know that 2⇒1, 3⇒4, and 2⇔3.

3⇒3: Recall Nakayama's lemma: if $R$ is a local ring with maximal ideal $\mathfrak{m}$ and $M$ is a finitely generated $R$-module, then any set of elements generating $M$ modulo $\mathfrak{m}M$ actually generates $M$. Applying this to $M = \mathfrak{m}$, we deduce that regular implies principal.

1⇒3: We have already shown that $\mathfrak{m}$ is invertible. Let $x \in \mathfrak{m}^{-1} \smallsetminus R$; we have seen in the proof of Proposition **??** that $x\mathfrak{m} \not\subset \mathfrak{m}$. Hence $x\mathfrak{m} = R \Rightarrow \mathfrak{m} = (x^{-1})$. □

## 5.4 Language: Dedekind rings, discrete valuation rings; and their properties

A ring as in Theorem **??** is called a *discrete valuation ring*. The name is because of the following equivalent characterization:

**Lemma 4.1.** *A domain is a discrete valuation ring if and only if there is a non-unit element $\varpi$ such that every nonzero element is of the form $x = u \cdot \varpi^n$ for some unit $u$ and natural number $n$. In this case, every ideal is of the form $(\varpi)^n$, and the map $v : x \mapsto n$ is a* valuation*: $v(xy) = v(x) + v(y)$ and $v(x+y) \geqslant \min(v(x), v(y))$.*

*Proof.* Let $R$ be a ring as in §**??**, and let $\varpi \in \mathfrak{m} \smallsetminus \mathfrak{m}^2$. Let $x \in R$ and let $n$ be the maximal integer such that $x\varpi^{-n} \in R$. Then $x\varpi^{-n} \notin \mathfrak{m}$ and hence it is a unit. The rest of the statements are left as exercises. □

A ring as in Proposition **??** is called a *Dedekind ring*. It has the following property:

**Proposition 4.2.** *In a Dedekind ring $R$, every non-zero ideal $I \subset R$ has a unique factorization $I = \prod_i \mathfrak{m}_i^{n_i}$ (finite product), where the $\mathfrak{m}_i$'s are prime (maximal) ideals and $n_i \in \mathbb{N}$. Every fractional ideal (=nonzero, finitely generated $R$-submodule of $K(R)$) has the same kind of an expression, with $n_i \in \mathbb{Z}$.*

*Proof.* Let $I$ be a maximal counterexample and $\mathfrak{m}$ some maximal ideal containing $I$, then $R \supset \mathfrak{m}^{-1}I \not\subset I$, again by the same integrality argument as in Lemma **??**. Therefore, by maximality, $\mathfrak{m}^{-1}I = \prod_i \mathfrak{m}_i^{n_i}$ and hence $I = \mathfrak{m} \cdot \prod_i \mathfrak{m}_i^{n_i}$. Uniqueness is left as an exercise. □

# Chapter 6

# Various notions of "smoothness"

*All rings in this lecture are noetherian.*

## 6.1 Normality and factoriality

A ring $R$ is called normal if and only if all local rings $R_{\mathfrak{p}}$ are integrally closed integral domains. This is equivalent to $R$ being a direct sum of (a finite number of, since noetherian) integrally closed integral domains. Notice that "direct sum" corresponds to disjoint union of the corresponding spectra. It is enough to check normality for $\mathfrak{p}$ =maximal ideals. For all these facts, see e.g. Ravi Vakil's section 5.4.1 (June 11, 2013 version).

**Theorem 1.1** (Serre)**.** *A ring is normal if and only if:*

(R1) *it is* regular in codimension one*, and*

(S2) *any prime associated to $(f)$, where $(f)$ is a non-zero-divisor and non-unit, is of height one; every associated prime of* $(0)$ *(i.e. of $R$) is of height zero.*

We explain the important condition (R1): it means that *for every prime* $\mathfrak{p}$ *of height less or equal to one*, the local ring $R_{\mathfrak{p}}$ is *regular*. For minimal primes (primes of height 0), this means that the local ring is a field. For primes of height one, it means that the local ring satisfies the equivalent conditions of Theorem **??** and hence is a discrete valuation ring. This should be thought of as saying that any singularities of the scheme are of codimension two or greater.

Regarding condition (S2), recall that by Krull's Hauptidealsatz all minimal primes over $(f)$ are of height one, anyway. Thus, the statement is that $R$ has no embedded primes, and neither does $R/(f)$ when $f$ is a non-zerodivisor.

*Remark.* The analogous conditions (R0): for every minimal prime the corresponding local ring is a field, and (S1): the ring has no embedded primes, characterize *reduced* rings. (Exercise!)

(S2) domains satisfy an analog of Hartogs' lemma for complex-valued functions in many variables (i.e. that they are holomorphic if they are holomorphic outside of codimension 2):

**Proposition 1.2** (Algebraic Hartogs' lemma)**.** *If $R$ is an (S2) domain (for example, a normal domain), then $R = \bigcap_{\mathfrak{p} \text{ of height } 1} R_{\mathfrak{p}}$.*

Let us understand what this says: the elements of $R_{\mathfrak{p}}$ are "rational functions", whose denominators are not identically zero on $\overline{\{\mathfrak{p}\}}$. In other words, there is a well-defined restriction of those functions to rational functions on $\overline{\{\mathfrak{p}\}}$ (i.e., a homomorphism: $R_{\mathfrak{p}} \to K(R)$. So, the statement is that if an element of $R$ does not have singularities along a codimension-one subscheme, then it is actually regular everywhere.

*Proof.* We claim: $x \in K(R)$ belongs to $R$ if and only if for every prime $\mathfrak{p}$ associated to a non-zerodivisor, $x$ belongs to $R_{\mathfrak{p}}$. If we prove this, we are done, because by the (S2) assumption these primes are all of height one.

Let $x = a/u \in K(R)$, then $u$ is a non-zerodivisor and we can assume that $a \notin (u)$. We claim that there is a prime $\mathfrak{p}$ associated to $(u)$ such that $a \notin \mathfrak{p}$. Indeed, since the ring $R' = R/(u)$ is noetherian there is a prime maximal among annihilators (in $R'$) of elements which contain $\operatorname{ann}_{R'}(a)$, and we have seen that all prime annihilators are associated primes. This proves the claim. □

A *factorial ring* or *unique factorization domain* is necessarily normal. We have the following equivalence (see Vakil, Proposition 11.3.5):

**Proposition 1.3.** *A noetherian domain is factorial iff all codimension/height 1 primes are principal.*

A scheme is called factorial if all local rings are factorial. This does *not* necessarily imply that it can be covered by schemes of the form $\operatorname{spec} R$ with $R$ factorial, see Vakil, 5.4N.

## 6.2 The Zariski cotangent space and regularity

Let $R$ be a local ring with maximal ideal $\mathfrak{m}$, $X = \operatorname{spec} R$. Then the $k(\mathfrak{m}) := R/\mathfrak{m}$-vector space $\mathfrak{m}/\mathfrak{m}^2$ is called the *Zariski cotangent space* of $X$ at $\mathfrak{m}$. Its linear dual is the *Zariski tangent space*, but as is always the case in algebraic geometry we first define notions through "functions", so the cotangent space is a more natural notion.

The reason that it is called the cotangent space is the following lemma. Recall that in differential geometry (over a field $k$) tangent vectors at a point $x$ can be thought of as derivations at that point, i.e. linear maps $D$ from the space of smooth functions to $k$ which satisfy the Leibniz rule:

$$D(f \cdot g) = f(x)D(g) + g(x)D(f).$$

**Lemma 2.1.** *There is a natural isomorphism between* $\mathrm{Hom}(\mathfrak{m}/\mathfrak{m}^2, k)$ *and the space of derivations:* $R \to k(\mathfrak{m})$.

The natural isomorphism of the lemma is sending a derivation to the linear functional defined by applying it to $\mathfrak{m}$ (indeed, it has to be zero on $\mathfrak{m}^2$); vice versa, given such a linear map we define the derivation as being equal to the functional on $\mathfrak{m}$, and equal to zero on constants. (Check that this indeed defines a derivation!.)

A local ring is called *regular* (or *nonsingular*) if the dimension of the Zariski cotangent space is equal to the dimension of the ring. This turns out to imply automatically that the local ring is integral.

(Examples discussed in class.)

A ring is called *regular* if all its local rings are regular. This is the best intrinsic notion of "smoothness" for rings. Serre proved that it is enough to check this for the localizations at maximal ideals only. This is certainly believable from a geometric point of view, but it is a hard theorem that actually requires a cohomological interpretation of regularity!

More precisely, Serre proved that a local ring is regular if and only if it is of *finite cohomological dimension*. We haven't learnt what this means, but keep it in mind, because it is an important notion; it shows that regularity is a deep property, and not just some explicit geometric feature.

## 6.3 Differentials and derivations

The definition of cotangent space at a point as $\mathfrak{m}/\mathfrak{m}^2$, or the tangent space as the space of derivations from $R$ to $k(\mathfrak{m})$, is not very satisfactory because it refers to a single point. We would like to define *sheaves* instead, and recover the tangent and cotangent spaces at points as fibers of these sheaves. In fact, we will define these sheaves in the relative setting: for a morphism $X \to Y$ we will define a relative tangent sheaf which will correspond to the "tangent spaces of the fibers" and a relative cotangent sheaf which will be the dual of this (hence a quotient of the cotangent sheaf by those differentials that vanish on "vertical tangent vectors").

Let $A$ be a ring and $R$ an $A$-algebra. An $A$-derivation $D : R \to R$ is an $A$-linear map which satisfies $D(fg) = fDg + gDf$. One can analogously define derivations $R \to M$, where $M$ is an $R$-module, which was the case above for $M = k(\mathfrak{m})$.

**Definition.** Let $f : X \to Y$ be a morphism of schemes. The *sheaf of* $\mathfrak{o}_Y$*-derivations on* $X$ is the sheaf (of $\mathfrak{o}_X$-modules) on $X$ associated to the presheaf on affine open sets:[1] $U \mapsto \mathrm{Der}_{f^{-1}(\mathfrak{o}_Y)(U)}(\mathfrak{o}_X(U), \mathfrak{o}_X(U))$.

The *sheaf or relative differentials* (with respect to the map $f$) is the sheaf $\Omega_{X/Y}$ (of $\mathfrak{o}_X$-modules) on $X$, together with an $f^{-1}(\mathfrak{o}_Y)$-derivation $d : \mathfrak{o}_X \to$

[1] Without the "affine" condition, this is not a presheaf, since a derivation on $\mathfrak{o}_X(U)$ does not necessarily determine its restriction to $\mathfrak{o}_X(V)$, when $V \subset U$; however, recall that a presheaf on affine open sets is enough in order to perform sheafification and define a sheaf.

$\Omega_{X/Y}$, which is universal (initial) among such objects, i.e. every other $f^{-1}(\mathfrak{o}_Y)$-derivation: $\mathfrak{o}_X \to \mathcal{M}$, where $\mathcal{M}$ is a $\mathfrak{o}_X$-module, factors through $d$ and a morphism of $\mathfrak{o}_X$-modules: $\Omega_{X/Y} \to \mathcal{M}$.

The definition of $\Omega_{X/Y}$ implies that, for any sheaf $\mathcal{M}$ of $\mathfrak{o}_X$-modules, we have $\mathrm{Der}_{f^{-1}(\mathfrak{o}_Y)}(\mathfrak{o}_X, \mathcal{M}) = \mathrm{Hom}_{\mathfrak{o}_X}(\Omega_{X/Y}, \mathcal{M})$. The sheaves defined above are quasi-coherent, and locally on $\operatorname{spec} R \to \operatorname{spec} A$ the sheaf $\Omega_{X/Y}$ corresponds to the module $I/I^2$, where $I \subset R$ is the kernel of the multiplication map: $R \otimes_A R \to R$ (it is generated by elements of the form $a \otimes 1 - 1 \otimes a, a \in A$).

More useful, in practice, is to take a closed immersion of $X$ (locally) into a $Y$-scheme $Z$, for instance $X \xhookrightarrow{j} Z = \mathbb{A}^n_Y$. Then, if $\mathcal{I}$ is the sheaf of ideals defining $X$, there is a short exact sequence:

$$\mathcal{I}/\mathcal{I}^2 \to j^*\Omega_{Z/Y} \to \Omega_{X/Y} \to 0. \tag{6.1}$$

For instance, if $Z = \mathbb{A}^n_Y$, and the ideal of $X$ is locally defined by a set of functions $f_i$, then $\Omega_{X/Y}$ is the quotient of the free $\mathfrak{o}_X$-module generated by symbols $dx_1, \ldots, dx_n$ by the submodule generated by the $df_i$'s (understood as elements of that module via the usual rules for computing differentials). See the discussion of the Jacobian criterion below.

For the composition of two maps: $X \xrightarrow{f} Y \to Z$ we have an exact sequence:

$$f^*\Omega_{Y/Z} \to \Omega_{X/Z} \to \Omega_{X/Y} \to 0. \tag{6.2}$$

We write $\Omega_X$ for derivations over $\operatorname{spec} \mathbb{Z}$. In particular, if $\mathfrak{m} \in X$ is a closed point with residue field $k(\mathfrak{m})$, then $\mathrm{Hom}_{\mathfrak{o}_X}(\Omega_X, k(\mathfrak{m})) = \mathrm{Der}_{\mathbb{Z}}(\mathfrak{o}_X, k(\mathfrak{m}))$ is the tangent space. Therefore, the fiber of $\Omega_X$ over $\mathfrak{m}$ is isomorphic to the cotangent space $\mathfrak{m}/\mathfrak{m}^2$.

*Remark.* For schemes defined over a field $k$, it is more natural geometrically to consider $k$-derivations instead of $\mathbb{Z}$-derivations. The two are not always equivalent: For example, consider schemes of the form $\operatorname{spec} E$, where $E$ is a field, containing another field $F$. A $\mathbb{Z}$-derivation of $\operatorname{spec} E$ is always trivial over the prime field $\mathbb{Q}$ or $\mathbb{F}_p$, but beyond that we have an inequality:

$$\dim_E \Omega_{\operatorname{spec} E/\operatorname{spec} F} \geqslant \mathrm{trdeg}_F E$$

for any finitely generated extension $E/F$ of $F$, with equality *if and only if* the extension is *separable* algebraic over a purely transcendental one. In particular, $\mathbb{Z}$-derivations and $F$-derivations are the same when $E$ is algebraic over $\mathbb{Q}$ or $\mathbb{F}_p$. However, if $E = k$ is transcendental over the prime field then it has non-trivial $\mathbb{Z}$-derivations, but of course no non-trivial $k$-derivations.

Finally, while we're at it, let's define the sheaf of differential operators. We will only consider the case when $X$ is defined over a field $k$, and the whole story will be $k$-linear. The sheaf of differential operators $\mathcal{D}_X$ is a sheaf of $\mathbb{N}$-filtered $\mathfrak{o}_X$-algebras of $k$-linear endomorphisms $D : \mathfrak{o}_X \to \mathfrak{o}_X$, with $\mathcal{D}^0_X = \mathfrak{o}_X$ and $\mathcal{D}^i_X$ consisting locally of those endomorphisms $D$ which satisfy:

$$[D, f] \in \mathcal{D}^{i-1}_X$$

for every $f \in \mathfrak{o}_X$ (considered as the operator of "multiplication by $f$"; and $[\bullet, \bullet]$ denotes the commutator).

It is easy to see that $\mathcal{D}^1_X = \mathfrak{o}_X + \mathrm{Der}_k(\mathfrak{o}_X, \mathfrak{o}_X)$. For nonsingular varieties the following is true: the sheaf of differential operators is generated (as an $\mathfrak{o}_X$-algebra) by derivations. This is not always true for singular varieties.[2]

## 6.4 Smooth morphisms

A morphism: $X \to Y$ of schemes is called *smooth* of relative dimension $r$ if, locally on $X$ it factors[3] as: $U \hookrightarrow \mathbb{A}^n_Y \to Y$ for some $N$ (where $U \subset X$ is open) so that locally on $U$ the ideal defining it is generated by $n - r$ sections $f_{r+1}, \ldots, f_n$, whose differentials $df_{r+1}(x), \ldots, df_n(x)$ are linearly independent in $\Omega_{\mathbb{A}^n_Y/Y} \otimes k(x)$, for every point $x$.

It is called *étale* if it is smooth of relative dimension 0. This turns out to be the correct algebro-geometric analog of the topological notion of a covering space.

The morphism is étale if and only if it is *flat* and *unramified*. "Recall" that an $A$-algebra $B$ is called flat if the functor $B \otimes_A \bullet$ from $A$-modules to $B$-modules is exact. "Unramified" means that the sheaf of relative differentials $\Omega_{X/Y}$ is zero and the morphism is locally of finite type. Equivalently, one can define "unramified" in a way similar to "smooth", except that we don't put a restriction on the number of sections $f_i$ generating the ideal, and we ask that the $df_i$ *generate* $\Omega_{\mathbb{A}^n_Y/Y} \otimes k(x)$. (The equivalence of the two definitions follows from (**??**).)

For example, closed immersions are unramified, but not etale (unless they are also open).

Here is a differential-geometric way to understand the notion of smoothness: We can consider the set of sections $(f_i)_{i=r+1}^n$ as a morphism $f : \mathbb{A}^n_Y \to \mathbb{A}^{n-r}_Y$. Such a morphism induces a pull-back of cotangent sheaves (over $Y$), and the differentials $df_i$ generate the image of $f^*\Omega_{\mathbb{A}^{n-r}_Y/Y}$ in $\Omega_{\mathbb{A}^n_Y/Y}$. Hence, saying that they are linearly independent at a point $x \in X$ is equivalent to saying that the corresponding map on tangent spaces:

$$T_x\mathbb{A}^n_Y \to T_0\mathbb{A}^{n-r}_Y$$

is surjective. (Well, we need to be precise about what "0" means: it should be replaced by the image of $x$, which is a point in the zero section of $Y$. Also, there is no "map on tangent spaces", in general, since $x$ may have a larger residue field than $f(x)$. In algebraic geometry what is really well-behaved is the pull-back of the sheaf of differentials.) The *implicit function theorem* in differential geometry would say that the fiber of 0 is smooth around $x$. But the fiber of 0 around $x$ is a neighborhood of $x$ in $X$!

[2] See http://cornellmath.wordpress.com/2007/09/09/d-module-basics-ii/.

[3] If the criterion is fulfilled for one such factorization, it is fulfilled for all.

Notice here that we have the exact sequence:

$$f^*\Omega_{\mathbb{A}^{n-r}_Y/Y} \to \Omega_{\mathbb{A}^n_Y/Y} \to \Omega_{\mathbb{A}^n_Y/\mathbb{A}^{n-r}_Y} \to 0,$$

and $\Omega_{X/Y}$ is (locally) the fiber of $\Omega_{\mathbb{A}^n_Y/\mathbb{A}^{n-r}_Y}$ over (a neighborhood of $x$ in) $X$; therefore, it is equal to the quotient of the free $\mathfrak{o}_X$-sheaf $\Omega_{\mathbb{A}^n_Y/Y}$ by the $\mathfrak{o}_X$-subscheaf generated by the $df_i$'s.

If $X, Y$ are smooth $S$-schemes, then $\Omega_{X/S}$ and $\Omega_{Y/S}$ are locally free, and an $S$-morphism $f : X \to Y$ of finite presentation is smooth (resp. étale) if and only if the map of differentials $f^*\Omega_{Y/S} \to \Omega_{X/S}$ is injective (resp. bijective). (Of course, from the exact sequence (**??**) and the last characterization of unramified morphisms it follows that "surjective" is equivalent to "unramified", without the assumption of smoothness.)

## 6.5 Jacobian criterion and $k$-smoothness

Let $k$ be a field, and $R$ a $k$-algebra of finite type. Hence, $R = k[x_1, \dots, x_n]/(f_1, \dots, f_m)$. Let $\mathfrak{m}$ be a $k$-valued point, i.e. a maximal ideal such that $R/\mathfrak{m} = k$. Notice that this is the same as a morphism: $\operatorname{spec} k \to \operatorname{spec} R$. Without loss of generality, by a change of coordinates, this point is the zero point. To compute the Zariski cotangent space we proceed as follows: First of all, for the ring $k[x_1, \dots, x_n]$ the Zariski cotangent space at zero is a vector space with basis $(dx_1, \dots, dx_n)$, where $dx_i$ denotes the image of $x_i$ modulo $(x_1, \dots, x_n)^2$. The cotangent space of $R$ at zero will be:

$$(x_1, \dots, x_n)/((x_1, \dots, x_n)^2, f_1, \dots, f_m).$$

Notice that $((x_1, \dots, x_n)^2, f_1, \dots, f_m)$ remains the same when we replace the $f_i$'s by their linear terms. Therefore, the Zariski cotangent space is generated by the $dx_i$'s modulo the relations which can be written as $df_i(0)$ (computed as differentials by the standard rules of calculus). This description holds at any $k$-valued point $x$, by replacing the point of evaluation 0 by $x$.

Another way to say that is that it is the quotient of the space with basis $(dx_i)_i$ by the image of the *Jacobian matrix:*

$$\operatorname{Jac}(x) := \left(\frac{\partial f_j}{\partial x_i}(x)\right)_{ij}.$$

We conclude:

> $\operatorname{spec} R$ is regular at a $k$-valued point $x$ if and only if the corank (dimension of cokernel) of the Jacobian matrix is equal to the dimension at $x$.

In particular, if $k$ is algebraically closed, then by Serre's theorem (i.e. that it suffices to check regularity only at closed points) this criterion on the Jacobian at every closed point is equivalent to regularity (because all closed points will be $k$-valued). This is part of a more general fact:

**Theorem 5.1.** *If a morphism $X \to \operatorname{spec} k$ is smooth, then $X$ is non-singular. The converse is true when $k$ is a perfect field.*

This is not true without the assumption of perfection. For instance, if $k = \mathbb{F}_p(t)$ then $k[x]/(x^p - t)$ is a field, hence regular, but the Jacobian criterion fails: if $f(x) = x^p - u$ then $df = 0$.

## 6.6 Formal smoothness and Hensel's lemma

`Later`

## 6.7 Examples from number theory

Let $E/F$ be a finite extension of $p$-adic fields (finite extensions of $\mathbb{Z}_p$), and let $\mathfrak{o}_E/\mathfrak{o}_F$ be the corresponding extension of rings of integers. Then $\mathfrak{o}_E = \mathfrak{o}_F[X]/f(X)$ for some polynomial $f$, which is a way to factor the morphism $\operatorname{spec} \mathfrak{o}_E \to \operatorname{spec} \mathfrak{o}_F$ through $\mathbb{A}^1_{\mathfrak{o}_E}$.

As we discussed above, the sheaf of relative differentials $\Omega_{E/F}$ is equal to the free sheaf generated by the symbol $dx$ by $df = f'(x)dx$. Thus, the morphism is unramified (actually, étale) if and only if $f'(x)$ is invertible in $\mathfrak{o}_F[X]/f(X)$, which is the case if and only if the extension is unramified.

Of course, $\operatorname{spec} E \to \operatorname{spec} F$ is always étale in this case. In positive characteristic, though, an finite extension of fields is étale if and only if it is separable. (If it is separable then it is generated by a single irreducible polynomial without multiple roots, and the above argument shows that it is etale. If it is inseparable, then it is ramified: the $E$-module $\Omega_{\operatorname{spec} E/\operatorname{spec} F}$ of relative differentials is nonzero. In general, an étale extension of (spec of) a field is, topologically, a union of closed points. Each of them is (spec of) a finite separable extension.

`Needs expansion`

# Chapter 7

# Divisors and line bundles

## 7.1 Weil divisors

In §**??** we introduced the notion of "regular in codimension one" (R1). This allows us to talk about Weil divisors and valuations.

Let $X$ be an integral scheme which is regular in codimension one. For every irreducible codimension-one $Y$ subscheme of $X$ (i.e. in the affine case $X = \operatorname{spec} R$, for any prime $\mathfrak{p}$ of $R$ of height one) we have a well-defined valuation homomorphism:

$$v_Y : K(X)^\times \to \mathbb{Z}.$$

Indeed, recall that the local ring $\mathcal{O}(X)_\mathfrak{p}$ is a valuation ring, so $v_Y$ denotes the corresponding valuation.

We let $\mathrm{Weil}(X)$ denote the free abelian group on the set of irreducible codimension-one subschemes of $X$ (i.e. height one primes of $R$). A Weil divisor is *effective* if the coefficients are all non-negative (sometimes written: $D \geqslant 0$).

We have a well-defined homomorphism:

$$K(X)^\times \to \mathrm{Div}(X)$$

by:

$$f \mapsto [f] := \sum_Y v_Y(f) \cdot Y.$$

The sum is actually finite, since every $f$ belongs only to finitely many primes of height one. (Indeed, these are the minimal primes over $(f)$.)

The image of $K(X)$ is called the subgroup*principal divisors* of $X$. The quotient of $\mathrm{Weil}(X)$ by principal divisors is the *class group* $\mathrm{Cl}(X)$ of $X$.

## 7.2 Cartier divisors and the Picard group

A cousin of Weil divisors are Cartier divisors: those are locally defined by rational functions modulo invertible regular functions. Let us assume that $X$ is

an integral (noetherian) scheme (i.e. reduced and irreducible), and consider the sheaf:

$$\mathcal{K}_X^\times/\mathcal{O}_X^\times$$

of invertible rational functions modulo invertible regular functions on $X$. A section of this sheaf is called a *Cartier divisor*, and we will denote the group of Cartier divisors by $\mathrm{Car}(X)$. A Cartier divisor is *effective* if it is locally generated by elements of $\mathcal{O}_X$, i.e. if it is an ideal sheaf.

Let us now further assume that $X$ is regular in codimension one. Then the discussion in the previous section gives a well-defined map:

$$\mathrm{Car}(X) \to \mathrm{Weil}(X).$$

This map is:

- injective, if $X$ is normal, since every rational function with zero divisor is invertible regular;
- bijective, if $X$ is factorial, since every codimension one prime in a factorial ring is principal. (Recall that "factorial" only refers to localizations at points of $X$, but a generator of the prime in the local ring will represent the divisor in a small neighborhood.)

The Picard group of a scheme $X$ is the abelian group of isomorphism classes of invertible sheaves (line bundles) on $X$. The multiplication operation is tensor product of line bundles. In this course we will see two notions of "Picard group": the first is set-theoretic, and the second is scheme-theoretic (under assumptions on $X$, the main one being properness). For now we focus on the set-theoretic one:

A *line bundle* on $X$ is an invertible sheaf of $\mathcal{O}_X$-modules on $X$, i.e. a rank-one locally free sheaf. Recall that in algebraic geometry we describe "spaces" by their functions, so it's natural to describe vector bundles by their sections. Their sections form a sheaf over $X$, and for example, in the differential-geometric category, a local isomorphism of a line bundle with $U \times \mathbb{C}$ (where $U$ is a small open set) gives rise to an isomorphism between smooth sections over $U$ and $C^\infty(U)$ – hence the property of sections being a locally free sheaf of rank one over the structure sheaf.

Tensoring line bundles corresponds to tensoring their sections over $\mathcal{O}_X$. Thus, isomorphism classes of line bundles form an abelian group, with the class of $\mathcal{O}_X$ being the identity element; the operation of taking inverse corresponds to replacing a sheaf $\mathcal{F}$ by $\mathrm{Hom}(\mathcal{F}, \mathcal{O}_X)$. The Picard group $\mathrm{Pic}(X)$, as a set, is the group of isomorphism classes of line bundles over $X$.

In the case of integral schemes, it turns out to be the same as the group of Cartier divisors modulo linear equivalence (i.e. modulo Cartier divisors arising from global rational functions).

**Lemma 2.1.** *If $X$ is an integral scheme, there is a natural isomorphism:* $\mathrm{Pic}(X) \simeq \mathrm{Car}(X)/K(X)^\times$.

*Proof.* The natural isomorphism is taking a line bundle $\mathcal{L}$, together with a rational section $s$ (there is always such a section, for example by trivializing over an open subset $U$ and taking the section corresponding to the constant $1 \in \mathcal{O}_X(U)$), to the Cartier divisor associated to this section. Different rational sections are obtained from each other by multiplication by an element of $K(X)^\times$, hence the map is well-defined.

Vice versa, given a class $D$ of Cartier divisors modulo rational equivalence, represented by a cover $(U_i)_i$ of $X$ and rational functions $f_i \in \mathcal{K}_X(U_i)$, we can define the subsheaf $\mathcal{O}_X(D)$ of $\mathcal{K}_X$ as the submodule of $\mathcal{K}_X$ generated by $f_i^{-1}$ over $U_i$. It is easy to see that this defines the inverse map.

By the way, if $X$ is normal then the construction of $\mathcal{O}_X(D)$ extends to any Weil divisor: locally, the sheaf consists of those rational functions $f$ such that $\div f + D \geqslant 0$. But it won't be invertible unless the divisor is also Cartier, i.e. locally principal (which, as we mentioned, is always the case when $X$ is factorial). □

*Example* 2.2. When $R$ is the ring of integers of a number field, or any Dedekind domain, the Picard group is the *ideal class group*, i.e. the quotient of the group of fractional ideals by the group of principal ideals. (Indeed, remember that since $R$ is a Dedekind domain, the set of fractional ideals form a free abelian group on the set of prime ideals – hence, it can be identified with $\mathrm{Pic}(\operatorname{spec} R)$.)

# Part II

# Basic notions: Representation Theory

# Chapter 8

# Representations of topological groups

*In this course, we will only consider representations on* ***complex*** *vector spaces; hence, we will be saying "vector space" and assuming implicitly that the underlying field is* $\mathbb{C}$. Representations over other fields, or even algebras, are very important as well, but we won't have time to discuss them. The interested student should read about the classification of semisimple algebras over an algebraic field (Wedderburn's theorem) in an algebra book (e.g. Lang's), and then follow this course trying to check where we are using the algebraic closedness and other properties of the field of complex numbers.

## 8.1 Definitions

A *representation* of a topological group $G$ on a topological vector space $V$ is a homomorphism

$$\rho : G \to \mathrm{GL}(V),$$

where $\mathrm{GL}(V)$ denotes the group of continuous, invertible linear operators on $V$, with the property that the induced "action" map:

$$G \times V \to V,$$

$$(g, v) \mapsto \rho(g)v$$

is continuous.

A *morphism* between representations $V_1$ and $V_2$ of $G$ is a continuous linear map: $V_1 \to V_2$ which commutes with the action of $G$. Hence, we have defined the *category of $G$-representations*.

A *subrepresentation* of $V$ is a *closed* subspace of $V$ which is stable under the action of $G$.

A representation is called *irreducible* if it does not contain any non-zero, proper subrepresentations. A finite-dimensional representation[1] is called (totally) *decomposable* if it is equal to the direct sum of irreducible subrepresentations.

*Example* 1.1. Consider the group $G = \mathbb{Z}$ and its representation $\rho$ on $V = \mathbb{C}^2$ with $\rho(1) = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$.

It contains an irreducible subspace $V_1$ generated by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, but it cannot be written as a direct sum of two irreducible subspaces. In other words, the following exact sequence of representations does not split:

$$0 \to V_1 \to V \to V/V_1 \to 0.$$

Given a decomposable representation $V$, it is isomorphic to a direct sum: $\oplus_i m(i)\pi_i$, where $m(i)\pi_i$ denotes the direct sum of $m(i)$ (assumed finite) copies of $\pi_i$ and the $\pi_i$'s are assumed to be non-isomorphic for different indices. Then $m(i)$ is called the *multiplicity* of $\pi_i$ and the subspace which is isomorphic to $m(i)\pi_i$ is called *the $\pi_i$-isotypic component of $V$*, sometimes denoted $V(\pi_i)$.

*Remark.* If our field of definition was not $\mathbb{C}$, but some non-algebraically closed field $k$, then we would distinguish between *irreducible* and *absolutely irreducible* representations: namely, $V$ is called absolutely irreducible if $V \otimes_k \bar{k}$ is irreducible. For example, the representation of the group $G = \mathbb{R}$ on the two-dimensional $\mathbb{R}$-space generated by the functions $\sin x$ and $\cos x$ is irreducible; but when we tensor with $\mathbb{C}$ it becomes reducible, namely the direct sum of the span of $e^{ix}$ and $e^{-ix}$.

A *unitary* representation of $G$ is a representation of $G$ on a Hilbert space $V$ which preserves the norm (i.e. $\rho$ has image in the subgroup of unitary transformations, $U(V) \subset \mathrm{GL}(V)$). This is the most important class of representations, because it has the nicest decomposition properties.

## 8.2 Examples; $G$-spaces and the regular representation

Most examples and applications of representation theory arise in the following, or similar, settings: a (topological) group acts on a (topological) space $X$, and then we get the *regular representation* of the group on various spaces of functions on $X$. In general, since our group is not always abelian, we will get used to letting the group act *on the right* on the space, so that its action on the function space is *on the left*:

$$(x, g) \mapsto x \cdot g,$$
$$(g \cdot f)(x) := f(x \cdot g). \tag{8.1}$$

[1] For infinite-dimensional representations the way of decomposing includes more than direct sums, and will be discussed later.

*Exercise.* Verify that (**??**) defines a left action.

Let us see some more examples:

*Example* 2.1. The action of the circle group $S^1$ on $L^2(S^1)$. We know by the theory of Fourier series that every $f \in L^2(S^1)$ can be decomposed uniquely as a convergent series:

$$f = \sum_{n=-\infty}^{+\infty} \hat{f}(n) z^n.$$

What is so special about the functions $z \mapsto z^n$? (If we identify $S^1$ with $\mathbb{R}/\mathbb{Z}$ via $x \mapsto e^{2\pi i x}$, these are the exponentials $e^{2\pi i n x}$.) The answer is that each of them spans an *irreducible representation* for the action of $S^1$, more precisely each one is an *eigenvector* for $S^1$, hence the representation that it spans is one-dimensional.

*Example* 2.2. The action of $\mathbb{R}$ on $L^2(\mathbb{R})$. Here we have the theory of Fourier transform, according to which every $f \in L^2(\mathbb{R})$ admits an essentially unique decomposition:

$$f = \int_{s \in i\mathbb{R}} \hat{f}(s) e^{2\pi s x} ds.$$

Again, the functions $x \mapsto e^{2\pi s x}$ are eigenvectors for the action of the group. We will recall/explain elsewhere what is the precise meaning of this "direct integral" decomposition.

*Example* 2.3. The group $D_8$ of symmetries of the square has eight elements. It acts on functions on the four vertices of the square. Can you decompose this four-dimensional space into irreducibles?

*Example* 2.4. Consider a positive definite quadratic form on $\mathbb{R}^3$. The corresponding special orthogonal group $G = \mathrm{SO}_3$ acts on $L^2(S^2)$. (The two-sphere $X = S^2$ is a *homogeneous* space for $G$ – i.e. $G$ acts transitively on its points. By choosing a point $v \in X$ we get an isomorphism: $X = \mathrm{SO}_2 \backslash \mathrm{SO}_3$, where $\mathrm{SO}_2$ is the special orthogonal group for the quadratic form restricted to the orthogonal complement of $v$.

The space $L^2(X)$ decomposes into an orthogonal direct sum of finite-dimensional, irreducible subrepresentations, and the elements of those subrepresentations are called *spherical harmonics*.

Finally, we can generalize the notion of the regular representation to line bundles: If $\mathcal{L}$ is a line bundle over a $G$-space $X$, endowed with a compatible $G$-action (such a line bundle is called *$G$-linear*), then the space of global sections $H^0(X, \mathcal{L})$ becomes a representation of $G$.

*Example* 2.5. Let $X = \mathbb{C}P^1$, $G = \mathrm{Aut}(X) \simeq \mathrm{PGL}_2(\mathbb{C})$, $\mathcal{L} = \mathcal{O}(n)$ for some integer $n$. Iff $n \geqslant 0$, the space of global sections of $\mathcal{L}$ is non-empty, and furnishes a representation of $G$. Explicitly, if we choose a point $x \in X$, its stabilizer $G_x$ is a *Borel subgroup* of $G$; we may fix an isomorphism $G \simeq \mathrm{PGL}_2(\mathbb{C})$ such that $G_x = B$ =the subgroup of upper triangular matrices (rather, their image in $\mathrm{PGL}_2$).

The space of global sections of $\mathcal{O}(n)$ can be identified with the space of homogeneous polynomial functions on $\mathbb{C}^2$ of degree $n$. It does not carry, in general, a natural action of $G$, but it carries a natural action of $\tilde{G} := \mathrm{GL}_2$, and can be identified with the space of regular functions on $\tilde{G}$ which have the property that for every $g \in G$:

$$f\left(\begin{pmatrix} a & * \\ & d \end{pmatrix} \cdot g\right) = a^n f(g).$$

*Exercise.* Can you verify these facts?

## 8.3 Discussion of the continuity condition

If $V$ is a Banach space, it seems tempting to endow $\mathrm{GL}(V)$ with the norm topology and to ask that $\rho$ is continuous with respect to that. However, this is *not* a good requirement, because many natural representations fail to satisfy it. For example, see the following exercise:

*Exercise.* We let the circle group $S^1 := \{z \in \mathbb{C} || z| = 1\}$ act on $X := \mathbb{R}^2$ by rotations in the obvious way. We consider the *regular representation* of $S^1$ on $V := L^p(X)$ $(p \geqslant 1)$:

$$(g \cdot f)(x) = f(g \cdot x).$$

Show that the resulting map: $G \to B(V)$ is not continuous. (Here $B(V)$ denotes the Banach space of bounded endomorphisms of $V$.)

Is the action map $G \times V \to V$ continuous? Including the case $p = \infty$?

# Chapter 9

# Representations without topology, discrete groups, finite-dimensional constructions

## 9.1 Representations on vector spaces without topology

We consider first the case that $V$ does not have topology. To be more precise: we ignore the topology on $V$ by giving it the weak topology induced by the set of *all* linear functionals. Open sets of this topology are generated by preimages of open sets in $\mathbb{C}$ under linear functionals; it is an easy exercise to show that then all linear endomorphisms of the vector space are continuous. In this case we have the following (easy!) theorem:

**Theorem 1.1.** *The category of representations of $G$ on vector spaces without topology is abelian.*

For the definition of an "abelian category", see the section on category theory in the appendix.

*Exercise.* Prove the theorem!

This has several implications (see the appendix on category theory), for instance:

**Corollary 1.2** (Morphisms between non-isomorphic representations)**.** *Let $\pi_1, \pi_2$ be non-isomorphic, irreducible representations of $G$ without topology. Then any morphism: $\pi_1 \to \pi_2$ is zero.*

**Corollary 1.3** (Jordan-Hölder theorem)**.** *If a representation $V$ admits a finite filtration by subrepresentations:*

$$0 = V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n = V$$

*where the consecutive quotients* $\operatorname{gr} V_i := V_i/V_{i-1}$ *are simple (i.e. irreducible and non-zero), then the (unordered, with multiplicities) set of isomorphism classes of the* $\operatorname{gr} V_i$*'s does not depend on the filtration chosen.*

A representation (without topology) is called *finitely generated* if there is a finite number of elements in $V$ such that $V$ is the smallest subrepresentation containing them. (In the case of topological representations, we usually apply the related notion of a *topologically finitely generated* representation, which is defined the same way but remembering that "subrepresentation" implies a *closed* subspace.)

*Exercise* (Important!). Prove that a representation of finite length is finitely generated.

*Example* 1.4. We consider the regular representation of the group $G = \mathbb{R}$ on the space of functions on $\mathbb{R}$. The subspace of functions of the form $P(x)e^{sx}$, where $s$ is a parameter and $P$ is a polynomial of degree $\leqslant N$, is a representation of finite length. (Prove!) What is its length?

*Example* 1.5. The same, essentially, example is the regular representation of $G = \mathbb{Z}$ and the subspace of functions of the form $P(n)t^n$ where $t \in \mathbb{C}^\times$ and $P(n)$ has degree $\leqslant N$.

*Example* 1.6. On the other hand, we claim that the regular representation of $G = \mathbb{Z}$ on $C_c(\mathbb{Z}) = \mathbb{C}[\mathbb{Z}]$ is *not* of finite length, although it is finitely generated! The fact that it is finitely generated is easy: it has a vector space basis consisting of the elements $\delta_n$ (delta function at the point $n$), and $\delta_n = (-n) \cdot \delta_0$, so the representation is generated by just the vector $\delta_0$.

There are two ways to prove that it is not of finite length: the first is to show that all irreducible representations of the group are characters one-dimensional (cf. Exercise **??**), and show that there are no irreducible subrepresentations in $C_c(\mathbb{Z})$. (Indeed, all eigenfunctions of the group on $C[\mathbb{Z}]$ are of the form $n \mapsto ct^n$, and this is easy to show.)

A second method is to show that $C_c(\mathbb{Z})$ has infinitely many, non-isomorphic, irreducible *quotients*. Then it cannot have a finite composition series, because by the Jordan-Hölder theorem this would imply that only a finite number of irreducible quotients, up to isomorphism, can appear.

Let $\chi : G \to \mathbb{C}^\times$ be the character: $n \mapsto t^n$ (for some $t \in \mathbb{C}^\times$). Clearly, different $t$'s give different homomorphisms, so there are infinitely many such. The map

$$T : f \mapsto \int_{\mathbb{Z}} f(n)\chi^{-1}(n)dn$$

(of course, the integral is a sum, but we write it so in order to get used to more general constructions) defines a morphism: $C_c(\mathbb{Z}) \to \mathbb{C}_\chi$, where $\mathbb{C}_\chi$ is the vector space $\mathbb{C}$ where $G$ acts via $\chi$.

Indeed, we have: $T(z \cdot f) = \int_{\mathbb{Z}} f(n+z)\chi^{-1}(n)dn = \chi(z)\int_{\mathbb{Z}} f(n)\chi^{-1}(n)dn = \chi(z)Tf$.

This morphism is clearly non-zero, since $T(\delta_0) = 1$. Therefore, $C_c(\mathbb{Z})$ has infinitely many non-isomorphic quotients and cannot be of finite length.

## 9.2 Discrete groups; the group algebra

Every time that we have a representation $\rho$ of a group, we naturally get a representation of its *group algebra* $\mathbb{C}[G]$ (consisting of formal, finite linear combinations of elements of $G$), i.e. a homomorphism of algebras:

$$\mathbb{C}[G] \to \mathrm{End}(V)$$

where, again, End stands for continuous endomorphisms. The action is defined as follows:

$$\rho(\sum_i c_i g_i) := \sum_i c_i \rho(g_i).$$

If the group is discrete, we have the following (easy!) theorem:

**Theorem 2.1** (When $G$ is discrete)**.** *The natural functor:*

$$\textit{representations of } G \longrightarrow \textit{topological } \mathbb{C}[G]\textit{-modules}$$

*is an equivalence of categories.*

*Exercise* (Very important!). Classify the isomorphism classes of finitely generated, without topology, representations of the group $\mathbb{Z}$. (Hint: describe the group algebra of $\mathbb{Z}$ and apply a well-known theorem from algebra.)

## 9.3 Finite dimensional representations – various constructions

When the vector spaces are finite-dimensional we can use various constructions from linear algebra in order to obtain more representations. (When the spaces are infinite-dimensional then "linear algebra" is called "functional analysis" and things are a bit more complicated.)

### 9.3.1 Direct sums

If $\pi_i, i = 1, \ldots, n$, are representations of $G$ then $\oplus_{i=1}^n \pi_i$ is a representation of $G$.

### 9.3.2 Tensor products

- If $\pi_i, i = 1, \ldots, n$, are representations of the groups $G_i$, respectively, then $\pi_1 \otimes \cdots \otimes \pi_n$ is a representation of $G_1 \times \cdots \times G_n$.

If the groups $G_i$ are the same ($= G$) then we often think of the tensor product as a representation of just one copy of $G$, by restricting the above to the diagonal copy of $G$.

- If $\pi$ is a representation of the group $G$, then $\pi \otimes \cdots \otimes \pi$ ($n$ times) is a representation of $G \times S_n$ (where $S_n$ is the symmetric group in $n$ elements).

  It contains the subrepresentation $S^n\pi$ of symmetric vectors in $\pi^{\otimes^n}$, and the subrepresentation $\wedge^n\pi$ of alternating vectors. This is a special case of the following phenomenon:

  **Lemma 3.1.** *Suppose that $W$ is a representation of $G_1 \times G_2$, and that restricting it to $G_2$ (forgetting the action of $G_1$) it is decomposable, then for any irreducible representation $\pi_2$ of $G_2$, the corresponding isotypic subspace $W(\pi_2)$ is $G_1 \times G_2$-stable.*

  *Remark.* For commuting (let's say invertible, though it's not necessary) linear operators (in other words, representations of $\mathbb{Z} \times \mathbb{Z}$) this is the well-known fact that for every eigenvalue of one, the corresponding eigenspace is stable under the action of the other.

  *Proof.* The fact that it is $G_2$-stable follows from the definition of "isotypic subspace". The action of every $g_1 \in G_1$ defines a $G_2$-morphism: $W(\pi) \to W$. If we write $W$ as a direct sum of its isotypic subspaces, and project to one of them (other than the $\pi$-th), we get a morphism: $W(\pi) \to W(\pi')$. By an easy extension of Corollary **??**, this has to be zero. In other words, $g_1 \cdot W(\pi) \subset W(\pi)$. □

  Assume now that $W$ is decomposable with respect to both $G_1$ and $G_2$.

  *Example* 3.2. The *alternating representation* or *sign character* sgn of the symmetric group $S_n$ (the group of permutations of $\{1, 2, \ldots, n\}$) is the character $S_n \to \mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^\times$, which sends any transposition to $-1$. It is well known that every element of $S_n$ can be written as an even or odd product of transpositions, but not both, hence this is well-defined. The space $\wedge^n V$ of alternating vectors in $V^{\otimes^n}$ is precisely the sgn-isotypic subspace with respect to teh $S_n$-action.

### 9.3.3 Inner products

Given a representation, can we find an inner product on it which is preserved by the action of the group? If we can, this is very advantageous:

**Proposition 3.3.** *Any finite-dimensional unitary representation is decomposable.*

*Proof.* Since $G$ preserves the inner product, the orthogonal complement of an invariant subspace is also invariant. By induction on the dimension, we see that the represenation decomposes into a direct sum of irreducibles. □

### 9.3.4 Duals

Given a representation $(\pi, V)$, we define a representation $(\pi^\vee, V^\vee)$ on the dual vector space by:

$$(g \cdot v^*)(v) = v^*(g^{-1}v). \tag{9.1}$$

Check that this indeed defines a representation (and that inverting $g$ was necessary)!

The definition looks better if we write with angular brackets, and it is equivalent to stating that:

$$\langle g \cdot v, g \cdot v^* \rangle = \langle v, v^* \rangle . \tag{9.2}$$

In other words: the action on the dual space is precisely that action under which the canonical pairing $V \otimes V^\vee \to \mathbb{C}$ is invariant.

For two representations $(\pi_1, V)$ and $(\pi_2, W)$, we can also consider the space $\mathrm{Hom}_{\mathbb{C}}(V, W)$, and define the natural action of $G \times G$ as a composition of operators:

$$((g_1, g_2) \cdot T)(v) = \pi_2(g_1) \cdot (T(\pi_1(g_2^{-1}) \cdot v)). \tag{9.3}$$

Then we have:

**Lemma 3.4.** *The natural map $n : W \otimes V^\vee \to \mathrm{Hom}_{\mathbb{C}}(V, W)$ is $G \times G$-equivariant, and an isomorphism if $V$ is finite-dimensional.*

*Proof.* Let us recall that the "natural map" is given by:

$$n(w \otimes v^*)(v) = \langle v, v^* \rangle \, w,$$

and it is a simple exercise in linear algebra that it is injective and (if the vector spaces are finite-dimensional, by dimension counting) surjective.

Let us check that it is equivariant:

$$n(g_1 \cdot w \otimes g_2 \cdot v^*)(v) = \langle v, g_2 \cdot v^* \rangle \, g_1 \cdot w = g_1 \cdot \big( \langle g_2^{-1} v, v^* \rangle \, w \big) = (g_1, g_2) \cdot n(w \otimes v^*)(v).$$

□

# Chapter 10

# Representations of finite groups

Throughout this lecture, all groups are finite.

## 10.1 Local finiteness

**Lemma 1.1.** *For any representation $V$ of a finite group $G$, any vector $v$ generates a finite-dimensional subrepresentation.*

*Proof.* The vector $v$ generates the subrepresentation $\overline{\mathbb{C}[G] \cdot v}$. Since $\mathbb{C}[G]$ is finite-dimensional, $\mathbb{C}[G] \cdot v$ is finite-dimensional and closed. □

A vector living in a finite dimensional subrepresenation of a representation is called a *finite* vector, and a representation consisting of finite vectors only is called *locally finite*.

Local finiteness allows us to focus on finite-dimensional representations only. In particular, we may mostly ignore the topology on our space – unless otherwise stated, all representations are without topology for this lecture.

**Corollary 1.2.** *Every irreducible representation of $G$ is finite-dimensional.*

## 10.2 Unitarity

**Lemma 2.1.** *Every representation of $G$ on a Hilbertian space is unitarizable.*

A Hilbertian space is a topological space which is isomorphic to a Hilbert space; in particular, finite-dimensional spaces are such. (The reason we say "Hilbertian" instead of "Hilbert" is that we don't need an inner product to be given to us.)

*Proof.* By averaging! Let $\langle\ ,\ \rangle$ be any positive definite inner product on the space (call the space $V$), then we define a new inner product:

$$\langle v, w\rangle' = \sum_{g\in G} \langle gv, gw\rangle\,.$$

This is easily seen to be positive definite and $G$-invariant. □

**Corollary 2.2.** *Every representation of $G$ on a Hilbertian space decomposes into a (topological, i.e. closure of the algebraic) direct sum of irreducibles.*

## 10.3 Schur's lemma

**Lemma 3.1** (Schur's lemma)**.** *If $V$ is an irreducible representation of $G$ then* $\mathrm{End}_G(V) = \mathbb{C}$ *(acting by scalar multiplication).*

As a corollary, irreducible representations of commutative groups are one-dimensional.

*Remark.* This is not the case with arbitrary groups and representations! For instance, Per Enflo constructed an operator on an infinite-dimensional Banach space which has no closed invariant subspaces. Invertible operators with this property exist, too.

*Proof.* Let $T \in \mathrm{End}_G(V)$, and let $\lambda$ be an eigenvalue for $T$. By replacing $T$ by $T - \lambda I$ we may assume that $\lambda = 0$, i.e. $\ker T \neq 0$. But $\ker T$ is $G$-invariant: if $Tv = 0$ then $T(g\cdot v) = g\cdot Tv = g\cdot 0 = 0$. By irreducibility, $\ker T = V$, therefore $T = 0$. (I.e. the original $T$ was the scalar $\lambda I$. □

## 10.4 The regular representation

Let $H$ be a finite group. In order to understand *all* irreducible representations of $H$ it is enough, it turns out, to look at the regular representation on the space of functions on $H$. More precisely, we consider the space $C(H)$ as a representation of the group $G := H \times H$ by left and right multiplication:

$$(h_1, h_2)\cdot f(x) = f(h_1^{-1} x h_2).$$

Notice that we could think of $C(H)$ as the group algebra of $H$. However, as we discussed previously we prefer to think of the group algebra as "measures", while now we want to decompose a space of "functions".

We can (and will) think of $C(H)$ as $L^2(H,\mu)$, where $\mu$ is any invariant measure; it is then easy to check that this is a *unitary* representation. We will take $\mu$ to be probability measure, i.e. $\frac{1}{|H|}$ times counting measure.

Our goal is to prove the following theorem:

**Theorem 4.1.** *There is a canonical, orthogonal direct sum decomposition of $H \times H$-representations:*

$$C(H) \simeq \oplus_\pi \pi \otimes \pi^*, \tag{10.1}$$

*where $\pi$ ranges over a set of representatives*[1] *for the isomorphism classes of irreducible representations of $H$.*

## 10.5 Matrix coefficients

For any representation $\pi$ of $H$, if $\pi^*$ denotes its dual, we have a *canonical, $H \times H$-equivariant map*:

$$\pi \otimes \pi^* \to C(H) \tag{10.2}$$
$$v \otimes v^* \mapsto f_{v,v^*}(x) = \langle v, \pi^*(x)v^* \rangle .$$

The function $f_{v,v^*}$ is called the *matrix coefficient* of $v$ and $v^*$.

We have:

**Lemma 5.1.** *If $\pi$ is irreducible, the matrix coefficient map (**??**) is injective.*

*Proof.* We will see later [ref] that $\pi \otimes \pi^*$ is an irreducible representation of $G = H \times H$. Hence, the map is either zero or injective. It is evidently not zero, hence it is injective. □

This is already a very important conclusion: we can find *every* irreducible representation of $H$ in the space $C(H)$! (Notice that under just the action of the *left* copy, $\pi \otimes \pi^*$ is just equal to a direct sum of $\dim \pi$ copies of $\pi$.)

## 10.6 Exhaustion of $C(H)$

By unitarity, we know a priori that $C(H)$ decomposes as an orthogonal direct sum of irreducible representations of $G = H \times H$. We will see in §**??** that irreducible representations of $G$ are precisely those of the form $\pi \otimes \tau$, where $\pi$ and $\tau$ are irreducible representations of $H$. Therefore, the following result suffices for the proof of Theorem **??**:

**Lemma 6.1.** *Let $\pi, \tau$ be irreducible representations of $H$ and assume that $M : \pi \otimes \tau \to C(H)$ is a non-zero (hence injective) morphism. Then $\tau \simeq \pi^*$, and we can choose this isomorphism so that $M$ is the matrix coefficient map.*

*Proof.* First notice (prove!) that the composition $L : \pi \otimes \tau \xrightarrow{M} C(H) \xrightarrow{\mathrm{ev}_1} \mathbb{C}$ is an $H^{\mathrm{diag}}$-invariant functional on $\pi \otimes \tau$. We claim that it is non-zero. Indeed, if $L(v \otimes w) = 0$ for every $v \in \pi, w \in \tau$ then $M = 0$ because $M(v \otimes w)(x) = M(v \otimes x \cdot w)(1) = L(v \otimes x \cdot w)$. (In other words, the reason is that the delta measure at 1 generates the dual space of $C(H)$ under the $G$-action.)

[1] Using Schur's lemma, prove the following: For any isomorphic, irreducible representations $\pi_1$ and $\pi_2$ of $G$, the $G \times G$-representations $\pi_1 \otimes \pi_1^*$ and $\pi_2 \otimes \pi_2^*$ are *canonically* isomorphic. Hence, the representation $\pi \otimes \pi^*$ does not depend on the choice of representative.

This non-zero $H$-invariant functional defines a non-zero morphism: $\tau \to \pi^*$, which by irreducibility has to be an isomorphism.

Finally, for the isomorphism $\phi : \tau \xrightarrow{\sim} \pi^*$ determined by $L$, we claim that $M$ is just the matrix coefficient map. Again, the reason is that the delta measure at the identity generates the dual space and hence evaluation at 1 determines a morphism completely. (See the section on induced representations later to understand this point.) Precisely, this means:

$$M(v \otimes w)(x) = L(v \otimes x \cdot w) = \langle v, \phi(x \cdot w)\rangle = \langle v, x \cdot \phi(w)\rangle = f_{v,w}(x).$$

□

Now Theorem **??** follows: we know that $C(H)$ has to decompose into an orthogonal direct sum of irreducibles of $G$, we know that the irreducibles that appear are precisely those of the form $\pi \otimes \pi^*$, and we know that the space $\mathrm{Hom}_G(\pi \otimes \pi^*, C(H))$ is one-dimensional, with a *canonical* element, for each $\pi$.

## 10.7 Corollaries

Let us number $\pi_1, \pi_2, \ldots, \pi_k$ the distinct isomorphism classes of irreducible representations of $H$, and by $d_i$ the dimension of $\pi_i$. The fact that there are finitely many is just a consequence of the following:

**Corollary 7.1.** *We have* $\sum_i d_i^2 = |H|$.

*Proof.* Indeed, $\dim C(H) = |H|$ and $\dim \pi_i \otimes \pi_i^* = d_i^2$. □

**Corollary 7.2.** *There are as many isomorphism classes of irreducible representations of $H$ as conjugacy classes in $H$.*

*Proof.* Let $H$ act on $C(H)$ as the diagonal of $G = H \times H$. Explicitly, this is the representation obtained by the conjugacy action of $H$ on itself. Thus, the space of invariants $C(H)^H$ has a basis consisting of the characteristic functions of conjugacy classes.

On the other hand,

$$C(H)^H = \left(\sum_{i=1}^{k} \pi \otimes \pi^*\right)^H = \sum_{i=1}^{k} (\pi \otimes \pi^*)^H = \sum_{i=1}^{k} \mathbb{C}.$$

The last step is by Schur's lemma.

We have found two bases[2] for the space $C(H)^H$; one is indexed by conjugacy classes and the other is indexed by irreducible representations; hence, they should be identical. □

We will see more corollaries right below, when we discuss characters.

[2] In the second case we have not yet described a basis, but just the lines on which the basis elements will live on. We will discuss distinguished basis elements when we discuss characters.

## 10.8 Characters

If we think of $C(H)$ as the group algebra $\mathbb{C}[H]$ of $H$, i.e. identify functions with measures by multiplying them by the counting measure. Then (exercise!) $C(H)^H$ is precisely the *center* of $\mathbb{C}[H]$. Hence, the last corollary and its proof provide two distinct bases for the center of $\mathbb{C}[H]$.

To be precise, we don't yet have a basis indexed by representations; all we know is that the space $\pi \otimes \pi^*$ has a one-dimensional line of $H$-invariant vectors. Here we will see that there is actually a *canonical* invariant element in $\pi \otimes \pi^*$ (considered as a subspace of $C(H)$), the *character* of $\pi$.

Given a finite dimensional (not necessarily irreducible) representation $\pi$ of $H$, we define a functional on $C(H)$ by:

$$\chi_\pi(\mu) = \operatorname{tr} \pi(\mu).$$

It is actually enough to describe this functional on point masses, so we consider $H$ as a subset of $\mathbb{C}[H]$ and set:

$$\chi_\pi(h) = \operatorname{tr} \pi(h).$$

**Lemma 8.1.** *$\chi_\pi$ is a non-zero, $H$-invariant (i.e. conjugation-invariant) function, and it lives in the image of the matrix coefficients for $\pi$.*

*Proof.* It is non-zero because $\chi_\pi(1) = d_\pi$ (dimension of $\pi$).

It is conjugation-invariant because the trace of an operator is conjugation-invariant.

If we choose dual bases $(e_i)_i, (e_i^*)_i$ for $\pi$ and $\pi^*$ then we have:

$$\operatorname{tr} \pi(h) = \sum_i \langle \pi(h) e_i, e_i^* \rangle = \sum_i f_{e_i, e_i^*}(h),$$

hence in the image of the matrix coefficient map. □

For irreducible $\pi$ this function, the *character* of $\pi$, is our distinguished basis element for $C(H)^H$ corresponding to $\pi$. We will see in the next paragraph that this is an orthonormal basis with respect to probability Haar measure on $H$.

Properties of the character:

- $\chi_{\pi_1 \oplus \pi_2} = \chi_{\pi_1} + \chi_{\pi_2}$.
- $\chi_{\pi_1 \otimes \pi_2} = \chi_{\pi_1} \cdot \chi_{\pi_2}$.
- $\chi_{\pi^*} = \overline{\chi_\pi}$.

All are easy if you use bases to compute the trace. E.g., for the third one let's choose dual bases $(e_i)_i$ of $\pi$ and $(e_i^*)_i$ of $\pi^*$, then if $A$ denotes the matrix for $\pi(g)$, the matrix for $\pi^*(g)$ is ${}^tA^{-1}$. Notice that the eigenvalues of that are the inverses of eigenvalues for $A$, but since $A^{|H|} = 1$ all those are roots of unity and satisfy: $\lambda^{-1} = \bar{\lambda}$. This shows the third property.

## 10.9 Orthogonality

Recall that we are using probability Haar measure on $H$, i.e. $\langle f_1, f_2\rangle = \frac{1}{|H|}\sum_{h\in H} f_1(h)\overline{f_2(h)}$.

**Proposition 9.1.** *If $\pi, \tau$ are irreducible representations of $H$ then:*

$$\langle \chi_\pi, \chi_\tau\rangle = \begin{cases} 1 & \text{if } \pi \simeq \tau \\ 0 & \text{otherwise.} \end{cases} \tag{10.3}$$

*Proof.* The fact that they are orthogonal if $\pi \neq \tau$ follows from the fact that the images of matrix coefficients for $\pi$ and $\tau$ are orthogonal, as we have already discussed. There remains to compute the $L^2$-norm of $\chi_\pi$:

$$\|\chi_\pi\|^2 = \frac{1}{|H|}\sum_{g\in H}\chi_\pi(g)\overline{\chi_\pi(g)} = \int_H \chi_\pi\overline{\chi_\pi}.$$

We use the fact that $\chi_\pi \cdot \overline{\chi_\pi} = \chi_{\pi\otimes\pi^*}$ in order to write this as:

$$\int_H \chi_{\pi\otimes\pi^*} = \langle \chi_{\pi\otimes\pi^*}, 1\rangle .$$

We claim that for any representation $\tau$ we have: $\langle \tau, 1\rangle = \dim \tau^H$. Indeed, there is a projection operator $p : \tau \to \tau^H$ defined by:

$$p(v) = \int_H \tau(g)v dg.$$

Its trace is equal to $\dim \tau^H$, but on the other hand it is equal to:

$$\int_H \operatorname{tr}\tau(g)dg = \int_H \chi_\tau = \langle \chi_\tau, 1\rangle .$$

topological vector Therefore, we have: $\|\chi_\pi\|^2 = \dim(\pi\otimes\pi^*)^H = 1$. □

These orthogonality relations are sometimes called "row orthogonality", referring to the character table (to be discussed) where each row corresponds to a different irreducible representation.

Using the characters as an orthonormal basis for invariant functions, we can express the inner product of the characteristic functions of two conjugacy classes $c$ and $d$ as follows:

$$\langle 1_c, 1_d\rangle = \sum_\pi \langle 1_c, \chi_\pi\rangle \langle \chi_\pi, 1_d\rangle = \sum_\pi \overline{\chi_\pi(c)}\chi_\pi(d).$$

Hence we get:

**Proposition 9.2.** *If $c, d$ denote two conjugacy classes in $H$, we have:*

$$\sum_\pi \chi_\pi(c)\overline{\chi_\pi(d)} = \begin{cases} \frac{|c|}{|H|}, & \text{if } c = d, \\ 0 & \text{otherwise.} \end{cases} \tag{10.4}$$

These relations are sometimes called "column orthogonality".

## 10.10 Examples: the character tables of $S_4, S_5, A_5$ etc.

To be discussed in class.

## 10.11 Irreducible representations of products of groups

We left for the end the proof of:

**Proposition 11.1.** *For a product $G \times H$ of finite groups, the irreducible representations are precisely those of the form $\pi \otimes \tau$, where $\pi$ is an irreducible representation of $G$ and $\tau$ is an irreducible representation of $H$.*

The proof will be based on *Burnside's theorem* which you can find in any standard algebra book, and (a version of) which says:

**Theorem 11.2.** *If $V$ is a simple module, finite-dimensional over an algebraically closed field $k$, for an algebra $R$, then the map $R \to \mathrm{End}_k(V)$ is surjective.*

Notice that the converse is also true: if $R \to \mathrm{End}_k(V)$ is surjective, then $V$ has to be simple, because it is a simple $\mathrm{End}_k(V)$-module.

Now, to prove the proposition, apply the theorem to $R = \mathbb{C}[G \times H] = \mathbb{C}[G] \otimes \mathbb{C}[H]$. First, $\pi \otimes \tau$ is irreducible because $\pi$ and $\tau$ are, and the map $\mathrm{End}_{\mathbb{C}}(\pi) \otimes \mathrm{End}_{\mathbb{C}}(\tau) \to \mathrm{End}_{\mathbb{C}}(\pi \otimes \tau)$ is surjective. (Easy exercise in tensor products!) Hence, $R \to \mathrm{End}_{\mathbb{C}}(\pi \otimes \tau)$ is surjective, and therefore $\pi \otimes \tau$ is a simple $R$-module.

Vice versa, given a simple $R$-module $W$, let $\tau$ be an irreducible representation of $H$ such that $\mathrm{Hom}_H(\tau, W) \neq 0$. How to show that $W \simeq \pi \otimes \tau$ for some irreducible $\pi$ of $G$? Well, if secretly we know it then notice that $\mathrm{Hom}_{\mathbb{C}}(\tau, W) \simeq \tau^* \otimes \pi \otimes \tau$ and $\mathrm{Hom}_H(\tau, W) = \mathrm{Hom}_{\mathbb{C}}(\tau, W)^H \simeq \pi \otimes (\tau^* \otimes \tau)^H \simeq \pi$.

Hence, set $\pi = \mathrm{Hom}_H(\tau, W)$; it carries an action of $G$ from its action on $W$. We have a natural morphism:

$$\pi \otimes \tau \to W, \tag{10.5}$$

simply by taking $L \otimes v$ to $L(v)$ (where $L \in \pi$ is considered as a morphism from $\tau$ to $W$, and $v \in \tau$). Since we assumed that $\mathrm{Hom}_H(\tau, W) \neq 0$, this morphism is not zero and hence it is surjective (by the irreducibility of $W$). We claim that it is also injective. Notice that no non-zero pure tensor, i.e. no non-zero element of the form $L \otimes v$, $L \in \pi, v \in \tau$, can be in the kernel of the map. Therefore, our goal is to reduce the problem to pure tensors.

Choose projectors $p_i \in \mathrm{End}_k(\tau)$ onto one-dimensional subspaces $\langle v_i \rangle$ such that $\sum_i p_i = \mathrm{Id}$. Since $\mathbb{C}[H] \to \mathrm{End}_k(\tau)$ is surjective, we can by abuse of notation consider the $p_i$'s as elements of $\mathbb{C}[H]$. Let $f \in \pi \otimes \tau$ be in the kernel

of (**??**); then $p_i(f)$ is also in the kernel, but $\mathfrak{p}_i(f)$ is a pure tensor, hence it has to be zero. On the other hand, $f = \mathrm{Id}(f) = \sum_i p_i(f)$, therefore $f = 0$. This proves the proposition.

# Chapter 11

# Representations of compact groups

The treatment of representations of finite groups of the previous section can be directly generalized to take care of compact groups. The only new difficulty is that representations will not be locally finite, and hence we need to use some functional analysis – the spectral theorem for compact operators – in order to produce finite-dimensional invariant subspaces.

For this lecture all groups are compact; it is known[1] that any compact group possesses a unique up to scalar left and right-invariant measure, called the *Haar measure*; we will denote it by $dg$ and normalize that to be a probability measure ($dg(G) = 1$).

## 11.1 Unitarity

**Lemma 1.1.** *Every representation of $G$ on a Hilbertian topological vector space is unitarizable.*

*Proof.* Same proof: take any positive definite hermitian form, and integrate it over the action of the group in order to make it invariant. (Notice that the continuity assumptions imply that for any fixed vectors $v, v'$ in the space the function $g \mapsto \langle gv, gv' \rangle$ is continuous, in particular integrable.) □

## 11.2 Spectral theorems

Before we continue, recall the following spectral theorems from functional analysis.

[1] Every locally compact group has a unique up to scalar left Haar measure and a unique up to scalar right Haar measure, but these may not coincide, e.g. in the group of upper triangular invertible $2 \times 2$ matrices.

Let $\mathcal{H}$ be a Hilbert space. The adjoint of a (bounded) operator $T$ on $H$ is the operator $T^*$ with $\langle T^*v, w\rangle = \langle v, Tw\rangle$. An operator $T$ is called self-adjoint if $T = T^*$. The idea of the spectral theorem is that the space $\mathcal{H}$ decomposes as an "integral" of "eigenspaces" of $T$. A familiar case of an integral of eigenspaces is when $\mathcal{H} = L^2(\mathbb{R})$, in which case the theory of Fourier transform says:

$$\mathcal{H} = \int_{i\mathbb{R}} \langle e^{sx}\rangle \, ds,$$

with the spaces $\langle e^{sx}\rangle$ being eigenspaces for all translation operators. (These are not self-adjoint, but here I'm just trying to explain the notion of an integral of eigenspaces.)

**Theorem 2.1** (Spectral theorem for self-adjoint operators)**.** *If $T$ is a self-adjoint operator on a Hilbert space $\mathcal{H}$ then there is a measure space $(X, \mathcal{B}, \mu)$, a measurable function $\lambda : X \to \mathbb{R}$ and a unitary isomorphism: $\mathcal{H} \simeq L^2(X, \mu)$ which carries the operator $T$ to "multiplication by $\lambda$".*

An operator on a Hilbert space is *compact* if it can be approximated in the operator norm by operators with finite-dimensional range. Equivalently, if it maps bounded sets to precompact sets (i.e. sets whose closure is compact).

**Theorem 2.2.** *Let $T$ be a compact self-adjoint operator on a Hilbert space, then there is a sequence of eigenvectors $v_n$ with (real) eigenvalues $0 \neq \lambda_n \to 0$ such that:*

$$\mathcal{H} = \ker T \oplus \bigoplus_n \langle v_n \rangle .$$

*an orthogonal topological direct sum (i.e. the closure of the algebraic direct sum).*

In particular, all eigenspaces with nonzero eigenvalues are finite-dimensional.

It will be our effort throughout to produce self-adjoint operators, and even compact ones, out of the action of the group. Before we get there, let us see an immediate corollary of the first spectral theorem, which generalizes results that were obvious for representations without topology:

**Lemma 2.3.** *If $\mathcal{H}, \mathcal{H}'$ are two irreducible unitary representations and $T \in \mathrm{Hom}_G(\mathcal{H}, \mathcal{H}')$ then $T$ is a scalar multiple of an isometry.*

*Proof.* It is enough to show that $T^*T$ and $TT^*$ are scalars. These are self-adjoint bounded operators, and applying the spectral theorem we see that $\mathcal{H}$ cannot be irreducible unless the space $X$ of the spectral theorem is an atom (i.e. every measurable subset satisfies $\mu(A) = 0$ or $\mu(X \smallsetminus A) = 0$. This implies that $\lambda$ is essentially constant on $X$, and hence these operators are scalars. □

This lemma is often called "Schur's lemma" for unitary representations, however we want to call "Schur's lemma" something stronger – namely, that every $G$-endomorphism of $\mathcal{H}$ is a scalar. We could prove this abstractly using functional analysis; however, we will see that for compact groups all irreducibles are finite-dimensional. Therefore, the proof that we saw for finite groups will work.

## 11.3 Convolution operators

Given a representation $(\pi, V)$ on a Fréchet space,[2] we define an action of *finite* measures on $G$ by:

$$\pi(\mu)v = \int_G \pi(g)v\mu(v). \tag{11.1}$$

This is a *vector valued integral*, and it makes sense as an element of $V$. (The completeness of $V$ is necessary for that.) There are generally different notions of strong and weak vector valued integrals for topological vector spaces, but for Hilbert spaces they coincide. The weak notion is that for every $w \in V^*$ we have:

$$\langle \pi(\mu)v, w\rangle = \int_G \langle \pi(g)v, w\rangle \, \mu(v).$$

For more information, see ... .

A *continuous (resp. smooth, $L^1$ etc) measure* on $G$ is a measure of the form $f \cdot dg$, where $f$ is a continuous (resp. smooth, $L^1$ etc.) function. Convolution of continuous, resp. smooth or $L^1$ measures is again continuous, resp. smooth or $L^1$.

The Banach space of finite complex measures on $G$ (with norm equal to the total mass of the absolute value of the measure, i.e. $\|\mu\| = |\mu|(G)$) will be denoted by $M(G)$. Notice that $L^1(G)dg \to M(G)$ is an isometric embedding. The following will clarify why representations are not continuous in the operator topology (because when points converge this is not the case for the associated delta measures):

**Lemma 3.1.** *The map: $M(G) \to \mathrm{End}(V)$ is continuous. (Remark: on the subspace $L^1(G)dg$, which is an algebra under convolution, this map is a homomorphism of algebras.)*

*Proof.* Proof for $V$ a Banach space (for simplicity). We have:

$$\|\pi(\mu)(v)\| = \left\| \int_G \pi(g)(v)\mu(g) \right\| \leqslant \int_G \|\pi(g)(v)\| \, |\mu|(g).$$

We claim that there exists a constant $C$ such that $\|\pi(g)\| \leqslant C$ for all $g \in G$. Indeed, from continuity of the action map $G \times V \to V$ at zero we deduce that there is a neighborhood $U$ of $1 \in G$ such that $U$ maps a ball in $V$ into another ball, i.e. $\|\pi(g)\| < c_1$ for all $g \in U$ and some constant $c_1$. Choosing a finite number of translates $g_i U$ which cover $G$, and assuming that $\|\pi(g_i)\| < c_2$ for some $c_2$ and all $i$, the claim follows with $C = c_1 c_2$.

Hence:

$$\|\pi(\mu)(v)\| \leqslant C\|v\||\mu|(G),$$

which proves the claim. □

[2] A Fréchet space is a Hausdorff topological vector space, whose topology is defined by a countable family of semi-norms, and is complete with respect to the corresponding uniform (metric) structure. It is a generalization of Hilbert and Banach spaces which in many cases provides the correct setup for representation theory. An example of a Fréchet space which is not Banach is $C^\infty(X)$, $X$ a manifold.

## 11.4 Peter–Weyl theorems

Now let $H$ be a compact group, and we consider now the space $\mathcal{H} = L^2(H)$. It is a unitary representation for $G = H \times H$. As before, we have:

**Lemma 4.1.** *1. For any finite-dimensional irreducible representation $\pi$ of $H$, we have an embedding: $\pi \otimes \pi^* \to C(H) \subset L^2(H)$ given by the matrix coefficient map.*

*2. For non-isomorphic $\pi$'s, the images of these embeddings are orthogonal.*

Notice that the image of the matrix coefficient map consists of *(left and right) finite* vectors (since $\pi \otimes \pi^*$ is finite dimensional). Our goal is to prove:

**Theorem 4.2** (Peter–Weyl theorem)**.** *There is a canonical isomorphism:*

$$L^2(H) \simeq \bigoplus_{\pi} \pi \otimes \pi^*$$

*(Hilbert space direct sum, i.e. orthogonal and closure of the algebraic one), where $\pi$ runs over representatives for the isomorphism classes of irreducible, finite dimensional representations of $H$.*

And, more generally:

**Theorem 4.3.** *Every Fréchet representation of $H$ contains a dense subspace of finite vectors. In particular, every irreducible represenation is finite dimensional. Every Hilbert representation of $H$ is the Hilbert space direct sum of irreducibles.*

## 11.5 Compactness of convolution by continuous measures

The main tool to prove the above theorems is the following:

**Lemma 5.1.** *Let $L$ (resp. $R$) denote the left (resp. right) regular representation of $H$ on $L^2(H)$. For every continuous measure $\mu$ on $H$ the operator $L(\mu)$ (resp. $R(\mu)$) is Hilbert-Schmidt and, hence, compact.*

*Proof.* Let $\mu = hdg$. Then the operator $L(\mu)$ has the integral expression:

$$L(\mu)(f)(x) = \int_H K_h(x,y) f(y) dy,$$

where the kernel is given by:

$$K_h(x,y) = h(xy^{-1}) dh.$$

It is known that the Hilbert Schmidt norm of an integral operator $T$ with kernel $K$ on a measure space $(X, dx)$ is given by:

$$\|T\|_{HS}^2 = \|K\|_{L^2(X\times X)}^2.$$

(Indeed, recall that we defined Hilbert-Schmidt operators for a Hilbert space $\mathcal{H}$ as the image of the injection: $\mathcal{H}\otimes\overline{\mathcal{H}} \to \mathrm{End}(\mathcal{H})$; use this to prove this statement about the norm of an integral operator.)

In particular, $L(\mu)$ is Hilbert-Schmidt. (It was important here that the group was compact for the $L^2$-norm of $K$ to be finite.) Thus it is compact. (Again, it is an easy exercise using the explicit expression $\|T\|_{HS}^2 = \sum_i \|Te_i\|^2$ to show that a Hilbert-Schmidt operator maps bounded sequences to precompact sequences, hence is compact.) □

## 11.6 Proof of the main theorems

We prove the theorems of §**??** in steps. Intermediate bullets with larger indent are not needed, but have been added as remarks. The proofs of some steps are left as exercises.

- Let $\mu_n$ be a sequence of approximations of the identity, i.e. positive probability measures whose mass is eventually concentrated in any neighborhood of the identity. Then for any Banach representation $(\pi, V)$ and vector $v \in V$ we have $\pi(\mu_n)v \to v$.
  - For any subrepresentation $V$ of $L^2(\mathcal{H})$, continuous (even smooth) functions are dense in $V$.

    Indeed, it is enough to choose a continuous (resp. smooth) approximation of the identity and convolve elements of $V$ with it; the convolution of any function with a continuous (resp. smooth) measure is continuous (resp. smooth).
- Left-finite (or right-finite) functions are dense in $L^2(H)$.

  This is the most important step of the proof. Assume to the contrary that there is a non-zero closed subspace $V$ without any left finite functions. We can find a continuous, self-adjoint measure $\mu$ on $H$ such that $R(\mu)V \neq 0$. (Indeed, we can do this by approximating the identity by continuous, self-adjoint measures; here by self-adjoint we mean that the operator $R(\mu)$ is self-adjoint, which is equivalent to $h(g^{-1}) = \overline{h(g)}$ if $\mu = hdg$ – exercise!.) We have already proven that $R(\mu)$ is compact, hence by the spectral theorem there is a non-zero (real) eigenvalue $\lambda$ of $R(\mu)$, and the $\lambda$-eigenspace is finite-dimensional. But the $\lambda$-eigenspace for $R(\mu)$ is stable under the left action of $H$, hence there are left-finite vectors, a contradiction!
- Any Fréchet representation of $H$ contains a dense subspace of finite vectors.

  Indeed, from the previous assertion we know that we can find approximations of the identity by left-finite measures. But $\pi(\mu)v$ is finite if $\mu$ is left-finite.

  To be precise, since "approximations of the identity" were defined measure-theoretically, and an approximation in $L^2$ preserves neither the positivity

nor the support condition, here is what we mean: Choose an approximation of the identity by $L^2$-measures $f_n dg$. We can approximate $f_n$ in $L^2$, and hence in $L^1$, by left-finite functions, so take finite functions $h_n$ such that $\|h_n - f_n\|_{L^1(H)} < \frac{1}{n}$. Then – easy exercise – for any vector $v \in V$ we have: $\pi(h_n dg)v \to v$.

– We have a sequence of inclusions:

$$L^2(H)_{\text{fin}} \subset C^\infty(H) \subset C(H) \subset L^2(H), \tag{11.2}$$

where $_{\text{fin}}$ denotes left *and* right finite functions, and the finite functions are dense in the natural topology of any of the other spaces.

Indeed, we apply the previous statement to any of these Fréchet spaces, viewed as a represenation of the group $H \times H$. What may not be immediately clear is that all finite vectors are contained in $C^\infty$, but recall that any subrepresentation contains a dense subspace of smooth functions; thus, any finite-dimensional invariant subspace has to consist of smooth functions. By the way, left-invariant (or right-invariant) would be enough for smoothness, by this argument. But we will see that left-finite implies right-finite.

The rest of the steps for proving the Peter-Weyl theorem are precisely as in the finite group case. A posteriori, any left-finite vector has to be right-finite, as well, since it has to be contained in the sum of a finite number of left-isotypic components.

## 11.7 Characters

Characters of finite-dimensional representations are defined exactly as in the case of finite groups, and with respect to probability Haar measure they form an orthonormal basis of class functions on the groups. (Notice that here we do not have that the eigenvalues of $\pi(g)$ are roots of unity; nontheless they lie on the unit circle because otherwise the closure of the sequence $\pi(g^n)$ would not be compact. Therefore, we still have that $\chi_{\pi *} = \overline{\chi_\pi}$.)

## 11.8 Examples

For explicit constructions of the irreducible representations of the compact groups $SU(2), SO(3), U(2)$ and $O(3)$ without using any Lie theory, cf. section II.5 in Bröcker and Tom Dieck, *Representations of Compact Lie Groups*.

# Chapter 12

# Algebraic groups and Lie groups

We now embark on the study of representations of compact, and non-compact, Lie groups. Actually, we will focus on the structure and representation theory of *reductive algebraic groups* over $\mathbb{R}$, and later over the $p$-adic numbers. "Algebraic" means that they have the structure of an algebraic variety (compatibly with group operations), or equivalently that they can be embedded in some $\mathrm{GL}_n$ as a closed subgroup defined by polynomial equations. "Reductive" will be explained later. All compact Lie groups are reductive algebraic, and most of the interesting non-compact Lie groups are such.

The study of *continuous representations* of *compact Lie groups* goes in parallel with the study of *algebraic representations* of their *complexifications*, and with *finite-dimensional representations* of their *Lie algebras*. We will introduce these topics a little more generally, in order to be able to use them later for *non-compact Lie (algebraic) groups* and their *infinite-dimensional representations*.

## 12.1 Lie groups, group schemes, algebraic groups

A *Lie group* is a group in the category of differentiable manifolds. A Lie group is automatically real-analytic.

A *group scheme* (over a base scheme $S$) is a group in the category of ($S$-)schemes. If $S = \operatorname{spec} k$, where $k$ is a field in *characteristic zero*, then it is automatically *smooth* over $k$. This is not the case in positive characteristic: for example, consider the (smooth) additive group scheme over $k = \mathbb{F}_p$:

$$\mathbb{G}_a = \operatorname{spec} k[T]$$

with the obvious group structure. For instance, addition $\mathbb{G}_a \times \mathbb{G}_a \to \mathbb{G}_a$ is given by the morphism induced by:

$$k[T] \ni f(T) \mapsto f(T_1, T_2) \in k[T] \otimes_k k[T] = k[T_1, T_2].$$

Now consider the "Frobenius" homomorphism:

$$\begin{array}{rcl} \mathbb{G}_a & \to & \mathbb{G}_a \\ k[T] \ni f(T^p) & \leftarrow\!\shortmid & f(T) \in k[T]. \end{array}$$

The *kernel* $K$ of this homomorphism is, as a scheme, isomorphic to $k[T]/(T^p)$, with the embedding $K \to \mathbb{G}_a$ given by the quotient map:

$$k[T] \to k[T]/(T^p)$$

and the inherited addition morphism:

$$k[T]/(T^p) \ni f(T) \mapsto f(T_1, T_2) \in k[T]/(T^p) \otimes_k k[T]/(T^p) = k[T_1, T_2]/(T_1^p, T_2^p).$$

Notice that this is a $k$-group scheme with *a unique closed point* (the identity), but it is *not* the trivial $k$-group scheme $\operatorname{spec} k$, as it has non-trivial tangent space (=Lie algebra), i.e. it is not reduced (hence not smooth).

Other examples of group schemes that are not smooth can be obtained, e.g. over $\mathbb{Z}_p$, for instance by taking the subgroup of $\mathrm{GL}_2$ (defined over $\mathbb{Z}$) which stabilizes the quadratic form given by the diagonal matrix with entries $(p, p)$. The fiber of this over the generic point $\operatorname{spec} \mathbb{Q}$ is an orthogonal group in two variables (hence of dimension 1), while the fiber over the special point $\operatorname{spec} \mathbb{F}_p$ is $\mathrm{GL}_2$ (of dimension 4) – in particular, this is not a smooth group scheme.

In any case, an *algebraic group* over a field $k$ is a *smooth group scheme* over $k$, and if it is affine then it is called a *linear algebraic group.*

## 12.2 Extension and restriction of scalars

If $S' \to S$ is a morphism of schemes, and $X$ is an $S$-scheme, then the *extension of scalars from $S$ to $S'$* of $X$ is defined as the fiber product $X_{S'} = X \times_S S'$. For example, if all are affine, with $S = \operatorname{spec} A$, $S' = \operatorname{spec} B$, and $A \to B$ a morphism of algebras, then $X_{S'} = \operatorname{spec}(A[X] \otimes_A B)$.

Vice versa, there are circumstances when we can consider $S'$-schemes to be $S$-schemes. It will suffice for us to consider the case when $S = \operatorname{spec} A$ (assumed noetherian) and $S' = \operatorname{spec} B$ with $B$ an $A$-algebra which is a projective (hence, locally free), finite $A$-module, and $X$ is affine of finite type over $\operatorname{spec} B$.

The *restriction of scalars from $B$ to $A$* of $X$ is an (affine) $A$-scheme $\operatorname{Res}_{B/A} X$ with the universal property that for any algebra $R$ over $A$ we have:

$$\operatorname{Res}_{B/A} X(R) = X(R \otimes_A B).$$

By *Yoneda's lemma* in category theory, this property uniquely identifies $\operatorname{Res}_{B/A} X$, if it exists. Under the above hypothesis, it exists and can be described as follows (when $B/A$ is free, otherwise the same construction locally): Let $B[x_i]_i/(f_m)_m$ be a finite presentation of $R$. Let:

$$f_m((\sum_j y_{ij} e_j)_i) = \sum f_{mj} e_j,$$

where $f_{mj} \in A[y_{ij}]$. Then:

$$\mathrm{Res}_{B/A}\, X = \mathrm{spec}\, A[y_{ij}]_{i,j}/(f_{mj})_{mj}.$$

*Example* 2.1. Let $h$ be a non-degenerate hermitian form in $n$ variables. The subgroup $U_n$ of $\mathrm{GL}_n(\mathbb{C})$ stabilizing $h$ is not an algebraic subgroup over $\mathbb{C}$, because the equations used to define it involve complex conjugation. However, if we consider $\mathrm{GL}_n/\mathbb{C}$ as an algebraic subgroup over $\mathbb{R}$, by restriction of scalars, then $U_n$ *is* an algebraic $\mathbb{R}$-subgroup.

Moreover, the extension of scalars from $\mathbb{R}$ to $\mathbb{C}$ of $U_n$ is isomorphic to $\mathrm{GL}_n(\mathbb{C})$, i.e. $U_n \times_{\mathrm{spec}\,\mathbb{R}} \mathrm{spec}\,\mathbb{C} \simeq \mathrm{GL}_n/\mathbb{C}$ (exercise!).

## 12.3 From smooth schemes to smooth manifolds

**Proposition 3.1.** *The map $X \to X(\mathbb{R})$ is a functor from the category of smooth schemes over $\mathbb{R}$ to smooth manifolds.*

*Proof.* Recall[1] that an $\mathbb{R}$-scheme $X$ is called *smooth, of relative dimension $r$* if locally on $X$ there is an embedding: $X \supset U \hookrightarrow \mathbb{A}^n_{\mathbb{R}}$ such that:

- locally around any point $y$ in (the image of) $U$, the ideal defining $U$ as a subscheme of $\mathbb{A}^n_{\mathbb{R}}$ is generated by $(n-r)$ polynomials $g_{r+1}, \ldots, g_n$,
- the differentials $dg_{r+1}(y), \ldots, dg_n(y)$ are linearly independent over $\mathbb{R}(y)$.

Notice that the $g_i$'s define a map: $\mathbb{A}^n_{\mathbb{R}} \to \mathbb{A}^{n-r}_{\mathbb{R}}$, and that a neighborhood of $y$ in $U$ coincides with a neighborhood of $y$ in the preimage of 0. The condition on differentials is equivalent to saying that all points in that neighborhood of $y$ in $U$ are *regular*, i.e. the differential of this map is surjective on tangent spaces. If we apply this to $\mathbb{R}$-points, by the implicit function theorem this implies that the preimage of 0 is a smooth manifold.

It is easy to show that the smooth structure does not depend on the embedding chosen, and that the map $X(\mathbb{R}) \to Y(\mathbb{R})$ obtained by a morphism $X \to Y$ is differentiable. Thus, the functor $X \to X(\mathbb{R})$ is a functor into the category of smooth manifolds. □

## 12.4 Open and closed subgroups of Lie groups

For any Lie group $G$ we will be denoting by $G^0$ the connected component of the identity. It is a normal subgroup. (Indeed, it is path connected since it is a manifold, and a path from the identity to any point is carried over by conjugation to another path from the identity.)

**Proposition 4.1.** *Any open subgroup of $G$ contains $G^0$.*

[1] This has actually not been included yet in the algebraic part of the notes

*Proof.* Let $H$ be an open subgroup. Its complement is a union of (left, let's say) $H$-cosets, and since right multiplication takes open sets to open sets, those cosets are open. Hence, the complement of $H$ is open, therefore $H$ is both open and closed, and therefore it contains the connected component of the identity. □

It is not true that every subgroup of a Lie group is closed. For instance, any one-parameter subgroup in the torus $(\mathbb{R}/\mathbb{Z})^2$ with non-rational slope is dense, but not closed.

On the other hand, every closed subgroup is a Lie subgroup:

**Theorem 4.2** (Cartan)**.** *Every closed subgroup of a Lie group is a smooth manifold, hence a Lie subgroup.*

We will prove this in the next lecture, because it makes use of Lie algebras.

## 12.5 Compact Lie groups are algebraic

An amazing fact is that the passage from real algebraic groups to Lie groups also works the other way in the case of compact Lie groups: they can all be realized as the points of a real algebraic group. This was proven by Weyl, and the following strengthening is due to Chevalley. We will not define the new terms here (they will come in later in the course), and we will only prove Weyl's weak version:

**Theorem 5.1.** *The functor $G \mapsto G(\mathbb{R})$ is an equivalence between: the category of $\mathbb{R}$-anisotropic reductive $\mathbb{R}$-groups whose connected components have $\mathbb{R}$-points, and the category of compact Lie groups. If $G$ is such an $\mathbb{R}$-group then $G^0(\mathbb{R}) = G(\mathbb{R})^0$. The $\mathbb{R}$-group $G$ is semisimple if and only if $G(\mathbb{R})$ has finite center, and in such cases $G^0$ is simply connected in the sense of algebraic groups if and only if $G(\mathbb{R})^0$ is simply connected in the sense of topology.*

We will only prove the statement that every compact Lie group is algebraic. This follows from the following two propositions, which have independent interest:

**Proposition 5.2.** *Every compact Lie group has a faithful (i.e. trivial kernel), finite-dimensional representation.*

*Proof.* Let $\pi_1, \pi_2, \dots$ be an enumeration of the irreducible representations of $G$. We already know from the Peter-Weyl theorem that they are finite-dimensional. For every $n$, let $G_n$ be the kernel of the map: $G \to \mathrm{GL}(\pi_1 \oplus \cdots \oplus \pi_n)$. Hence, we have a sequence of closed subgroups:

$$G = G_0 \supset G_1 \supset G_2 \supset \dots.$$

We claim that every such sequence terminates. Indeed, by Cartan's theorem, we know that all $G_n$ are Lie groups, therefore the dimension of $G_n$ has to stabilize after some $n$. But then the map $G_n^0 \hookrightarrow G_{n+1}^0$ is eventually the identity

(the image is both open[2] and closed), and by compactness each $G_n$ has a finite number of connected components, so the sequence has to terminate.

On the other hand, the intersection of the $G_i$'s is (again by Peter-Weyl) the kernel of the left regular representation of $G$ on $L^2(G)$, hence trivial. □

The second element is of invariant-theoretic nature. For this, let $G \to \mathrm{GL}(V)$ be a (complex), finite-dimensional representation of $G$ and consider it as a real representation by regarding $V$ as a real vector space. (This is the baby case of "restriction of scalars".) Accordingly, $\mathrm{GL}(V)$ is considered as an algebraic group over $\mathbb{R}$ (by restriction of scalars). Notice that the Zariski closure[3] of the image of $G$ is a real algebraic subgroup. We need to show that it coincides with $G$. One thing that $G$ and its Zariski closure have in common is the set of invariants on the polynomial ring $\mathbb{R}[V]$. Recall that the polynomial ring $\mathbb{R}[V]$ is (essentially, by definition) the symmetric algebra on the dual space $S^\bullet V^*$.

**Proposition 5.3.** *Each orbit of a compact group (not necessarily a Lie group!) on the vector space $V$ of a representation is (the real points of) a real algebraic subset. The image of $G \to \mathrm{GL}(V)$ is (the $\mathbb{R}$-points of) an algebraic subgroup.*

*Proof.* The second statement follows from the first because $\mathrm{GL}(V) \hookrightarrow \mathrm{End}(V) = V^* \otimes V$, and the image is the orbit of the identity transformation.

For the first we consider the map: $V \to V /\!\!/ G := \operatorname{spec} \mathbb{R}[V]^G$, and the induced map on $\mathbb{R}$-points: $V(\mathbb{R}) \to V /\!\!/ G(\mathbb{R})$. Clearly, the preimage of any point is a union of $G$-orbits. We claim:

> The preimage of every $\mathbb{R}$-point contains at most one $G$-orbit on $V(\mathbb{R})$.

This will be enough to prove the claim: Since the preimage is an algebraic variety over $\mathbb{R}$, it means that $G$-orbits are the $\mathbb{R}$-points of algebraic varieties (maybe empty, because the preimage of an $\mathbb{R}$-point does not need to contain any $\mathbb{R}$-points – for instance, consider the quotient of $\mathbb{C}^\times$ by the circle group).

To prove the claim we must show that if $Y_1, Y_2$ are two distinct $G$-orbits on $V(\mathbb{R})$, then there is a $G$-invariant polynomial which takes different values on $Y_1$ and $Y_2$ (i.e. the ring of invariant polynomials separates $G$-orbits).

Notice that $\mathbb{R}[V]$ is a locally finite representation of $G$ (this follows by its identification with $S^\bullet V^*$), and therefore by the Peter-Weyl theorems it is completely reducible. If we fix two points $y_1 \in Y_1$ and $y_2 \in Y_2$, then the integrals:

$$\int_G f(y_i \cdot g) dg$$

represent two $G$-invariant functionals $\ell_1, \ell_2$ on $\mathbb{R}[V]$. They obviously factor through restriction of polynomials to the compact subset $Y_1 \cup Y_2$, and *by the*

[2] See the section on Lie algebras: ...

[3] It is important here that we have restricted scalars to $\mathbb{R}$, because the Zariski closure depends on whether we consider $\mathrm{GL}(V)$ as a complex or as a real variety; for example, the Zariski closure of the circle group $S^1$ in $\mathbb{C}^\times$ is $S^1$ or $\mathbb{C}^\times$, according as $\mathbb{C}^\times$ is considered as a real or complex variety.

*Stone-Weierstrass theorem, the restriction of polynomials is dense in the space of continuous functions* on $Y_1 \cup Y_2$. Therefore, $\ell_1$ and $\ell_2$ are linearly independent, i.e. $\ell_2$ is non-zero on the kernel $W$ of $\ell_1$.

Hence, $\ell_2$ defines a $G$-invariant functional: $W \to \mathbb{C}$, and by complete reducibility this splits; in particular, *there is a $G$-invariant element $f \in W$ with $\ell_2(f) \neq 0$*. That is, there is a $G$-invariant polynomial on $V$ whose integral over $Y_1$ is zero and whose integral over $Y_2$ is non-zero. But this means that its value on $Y_1$ is zero and its value on $Y_2$ is non-zero, which is what we wanted to prove. □

*Remarks.* 1. A similar argument works to establish the following important result: Let $G$ be a reductive algebraic group over an algebraically closed field $k$ in characteristic zero. We have not defined "reductive", but in characteristic zero this is equivalent to the statement that every algebraic representation of $G$ is completely reducible. Let $X$ be an affine variety on which $G$ acts. Then the closed points of $X /\!\!/ G := \operatorname{spec} k[X]^G$ are in bijection with (Zariski) closed orbits of $G$ on $X$.

Here is the proof: Let $Y_1, Y_2$ be two closed orbits and consider the $G$-stable ideal $I \subset k[X]$ of regular functions vanishing on $Y_1$. Restriction to $Y_2$ gives a map: $I \to k[Y_2]$, and the image $I'$ has to be non-zero because otherwise $Y_2$ would be in the Zariski closure of $Y_1$. But since $Y_2$ is a Zariski-closed orbit, a non-zero ideal coincides with the whole ring, therefore the image $I'$ of $I$ contains constant functions. By reductivity, there is a $G$-invariant quotient of $I'$, hence a $G$-invariant quotient of $I$. By reductivity, again, $I$ has a $G$-invariant element whose image in $I'$ is non-zero. In other words, $Y_1$ and $Y_2$ are separated by $G$-invariant regular functions.

2. The last proposition is not true for non-compact groups. For instance, not only is the subgroup:

$$\left\{ \begin{pmatrix} 1 & & \\ x & t & \\ y & & t^\alpha \end{pmatrix} : x, y \in \mathbb{R}, t \in \mathbb{R}_+^\times \right\}$$

of $\mathrm{GL}_3(\mathbb{R})$ (where $\alpha$ is an irrational number) not an algebraic subgroup of $\mathrm{GL}_3$, but it is not isomorphic to (the $\mathbb{R}$-points of) *any* real algebraic group.[4]

[4] For details, cf. http://terrytao.wordpress.com/2011/06/25/two-small-facts-about-lie-groups/.

# Chapter 13

# Lie algebras

## 13.1 Definitions

A *Lie algebra* is a vector space $\mathfrak{g}$ with a bilinear, antisymmetric operation $[\bullet, \bullet] : \mathfrak{g} \otimes \mathfrak{g} \to \mathfrak{g}$ with the following property:

> The map $X \mapsto \mathrm{ad}(X) := [X, \bullet] \in \mathrm{End}(\mathfrak{g})$ is a *representation* of $\mathfrak{g}$, in the sense that $\mathrm{ad}X\mathrm{ad}Y - \mathrm{ad}Y\mathrm{ad}X = \mathrm{ad}([X, Y])$.

The last condition is nothing but the well-known *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0. \tag{13.1}$$

An associative algebra $A$ gives rise to a Lie algebra by setting $[X, Y] = XY - YX$. In that sense, the notion of a "representation" of the Lie algebra defined above is nothing but a Lie algebra homomorphism: $\mathfrak{g} \to \mathrm{End}(V)$, where $V$ is a vector space.

A Lie algebra is not an associative algebra; however, there is an initial object $U(\mathfrak{g})$ in the category of associative algebras $A$ with a homomorphism of Lie algebras: $\mathfrak{g} \to A$. In other words, $U(\mathfrak{g})$, together with the homomorphism $\mathfrak{g} \to U(\mathfrak{g})$ is defined by the universal property that any other homomorphism of Lie algebras $\mathfrak{g} \to A$ factors uniquely through $U(\mathfrak{g})$. It is very easy to construct $U(\mathfrak{g})$ and to see that it has this property: Simply take the quotient of the tensor algebra of $\mathfrak{g}$:

$$\bigoplus_{n \geqslant 0} \mathfrak{g}^{\otimes^n}$$

by the two-sided ideal generated by all elements of the form: $X \otimes Y - Y \otimes X - [X, Y]$, $X, Y \in \mathfrak{g}$.

## 13.2 Poincaré-Birkhoff-Witt

Now let $(X_i)_{i \in I}$ be a linearly ordered basis of $\mathfrak{g}$ as a vector space. (We won't need it, but we include the possibility that $\mathfrak{g}$ is infinite-dimensional, even uncountable-dimensional.)

**Theorem 2.1** (Poincaré-Birkhoff-Witt)**.** *The monomials of the form $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$, with $i_1 < i_2 < \cdots < i_k$, form a vector space basis for $U(\mathfrak{g})$.*

*Proof.*

- They generate: Indeed, if we didn't have the condition $i_1 < i_2 < \cdots < i_k$ then we would have a basis for the tensor algebra. Now take an element $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$ (no condition on the $i_j$'s) which does not belong to the span of the elements in the asserted basis, and such that: $a := \sum r_i$ is minimal with this property; for the given $a$, the first index $j$ such that $i_j > i_{j+1}$ is minimal. Then by applying the relation: $X_{i_j} \otimes X_{i_{j+1}} - X_{i_{j+1}} \otimes X_{i_j} = [X_{i_j}, X_{i_{j+1}}]$ we can easily reach a contradiction. Hence the asserted basis is, at least, a generating set.

- To show that the elements are indeed linearly independent is much more difficult. Think about it for a moment: how do we even know that $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$ is not zero? Couldn't there be manipulations using the formula $X \otimes Y - Y \otimes X - [X, Y]$ that would eventually reduce it to zero? How can we know without even knowing what $[X, Y]$ is? (i.e. without having any extra information on the structure of the Lie algebra).

  It turns out that the only thing we need to know is the Jacobi identity. We will use it to construct a representation $\rho$ of $\mathfrak{g}$ (equivalently: of $U(\mathfrak{g})$) on the free vector space $V$ generated by the monomials $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$. This representation will have the property that $Y = X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$, considered as an element of $U(\mathfrak{g})$, takes $1 \in V$ to $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$ (a posteriori, it is just left multiplication on $U(\mathfrak{g})$). In particular, the map $U(\mathfrak{g}) \ni Y \mapsto \rho(Y)(1) \in V$ is injective, which proves the theorem.

  To define the representation, it is enough by the universal property of $U(\mathfrak{g})$ to define it on $\mathfrak{g}$, and in fact just on the basis elements. For $Y = X_i$, let $[X_i, X_j] = \sum_k a_{ijk} X_k$ and use these coefficients to define the action of $\rho(Y)$ on $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k} \in V$ in the obvious way. It can get messy, but using the Jacobi identity one can show that this is a well-defined represenation of $\mathfrak{g}$.

  For a nicer proof, cf. Braverman and Gaitsgory, *Poincaré-Birkhoff-Witt theorem for quadratic algebras of Koszul type,* J. Algebra 181 (1996), no. 2, 315–328. It uses algebraic deformation theory, deforming $U(\mathfrak{g})$ to $S(\mathfrak{g})$ (the symmetric algebra of $\mathfrak{g}$, which, as we will see, is the associated graded of $U(\mathfrak{g})$), and deduces the desired result from a flatness property ($\Rightarrow$ the fibers of the deformation have the "same dimension"). A posteriori, this deformation is a standard way to deform a filtered ring to its associated graded – we may discuss it at some point.

□

We consider $U(\mathfrak{g})$ as a filtered algebra: $U(\mathfrak{g}) = \bigcup_{n \in \mathbb{N}} F_n$, $F_{n+1} \supset F_n$, where $F_n$ is generated by products of at most $n$ elements of $\mathfrak{g}$. Recall that the associated graded of a(n $\mathbb{N}$-)filtered algebra $A$ is the algebra $\operatorname{gr} A = \oplus_n F_n / F_{n-1}$ (with $F_{-1} = 0$).

**Corollary 2.2.** *We have a canonical isomorphism:* $\operatorname{gr} U(\mathfrak{g}) = S(\mathfrak{g})$*, the symmetric algebra of* $\mathfrak{g}$*. The algebra* $U(\mathfrak{g})$ *is noetherian.*

*Proof.* The PBW theorem shows that $F_n$ has a vector space basis consisting of $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$, with $i_1 < i_2 < \cdots < i_k$ and $\sum i_j \leqslant n$. The relation $XY - YX = [X, Y]$ shows that the commutator of two elements in $F_n$ lies in $F_{n-1}$ and hence $\operatorname{gr} U(\mathfrak{g})$ is commutative. By the universal property of the symmetric algebra, there is a unique homomorphism of algebras: $\mathcal{S}(\mathfrak{g}) \to U(\mathfrak{g})$ which is the identity from $\mathfrak{g} \subset S(\mathfrak{g})$ to $U(\mathfrak{g})$. But this map is a bijection on distinguished bases, therefore an isomorphism of algebras.

The noetherian property now follows from the noetherian property of $S(\mathfrak{g})$ by the following standard argument: if $J_1 \subset J_2 \subset \cdots$ is an increasing sequence of ideals, then so is $\operatorname{gr} J_1 \subset \operatorname{gr} J_2 \subset \cdots$, where $\operatorname{gr} J = \oplus_n (J \cap F_n / F_{n-1})$. Notice that the map $J \mapsto \operatorname{gr} J$ is not injective on ideals: two different ideals of $U(\mathfrak{g})$ can have the same image in its graded. However, the map *is* injective on chains, i.e. if $J_1 \subset J_2$ and their graded ideals coincide, then $J_1 = J_2$. From the noetherian property of $S(\mathfrak{g})$, the sequence of graded ideals stabilizes, therefore so does the original sequence. □

There are many corollaries of the PBW theorem, one of them being that if $\mathfrak{h} \subset \mathfrak{g}$ is a Lie subalgebra then $U(\mathfrak{g})$ is a free $U(\mathfrak{h})$-module, and hence the *induction* functor:

$$M \mapsto U(\mathfrak{g}) \otimes_{U(\mathfrak{h})} M$$

(where $M$ is an $\mathfrak{h}$-module) is exact. We will come back to these applications as they become relevant.

## 13.3 The Lie algebra of a Lie or algebraic group

Often the Lie algebra of a Lie or algebraic group is defined as the tangent space to the identity element. (The tangent space of the identity makes sense, and is a vector space over the base field, both differential-geometrically and algebro-geometrically.) This definition gives no information as to where the bracket operation comes from. A better definition is in terms of *vector fields* or *derivations*. Look in the algebraic part of the notes for a discussion of derivations, in general "vector fields" or "derivations" are sections of the tangent bundle. The sheaf of $k$-derivations on a $k$-scheme $X$ can be considered as a subsheaf of:

$$\mathscr{H}om_k(\mathfrak{o}_X, \mathfrak{o}_X),$$

which is a sheaf of associative algebras. Hence, there is a Lie bracket in $\mathscr{H}om(\mathfrak{o}_X, \mathfrak{o}_X)$, and it can be seen that the derivations form a Lie subalgebra, i.e. the commutator of two vector fields is again a vector field.

Therefore, we define the Lie algebra of an algebraic group $G$ over $k$ to be the space of *left invariant $k$-linear derivations* on $G$. In characteristic $p > 0$ it has an additional structure of what is called a *restricted Lie algebra*, i.e. a Lie algebra endomorphism $D \to D^p$ which satisfies certain natural axioms (will not

discuss here; just notice that the $p$-th power of a derivation, considered as a differential operator, is again a derivation).

## 13.4 Exponential map

Now we work differential-geometrically in the setting of a real Lie group. A *one parameter subgroup* is a homomorphism of Lie groups: $\gamma : \mathbb{R} \to G$. Its differential at zero gives rise to an element $\gamma'(0) \in T_eG = \mathfrak{g}$.

**Proposition 4.1.** *The map $\gamma \mapsto \gamma'(0)$ is a bijection between one-parameter subgroups and elements of the Lie algebra.*

*Proof.* Locally around any point $x$, any vector field is uniquely integrable (this is a basic result from ODEs), namely: if $\mathbf{v}$ is a vector field then there is an interval $(-\epsilon, \epsilon)$ and a curve $\gamma : (-\epsilon, \epsilon) \to G$ such that $\gamma(0) = x$ and $\gamma'(t) = \mathbf{v}(\gamma(t))$, and the germs of any two such around 0 coincide.

For a left-invariant vector field, we can use left translations by the group to show that this local existence and uniqueness statement becomes global. □

This allows us to define an *exponential map*:

$$\mathfrak{g} \to G$$

by:

$$\exp(X) = \gamma_X(1),$$

where $\gamma_X$ is the unique one-parameter subgroup with $\gamma'(0) = X$.

This is not a group homomorphism, except if $G$ is abelian (but, by definition, it is a group homomorphism when restricted to any one-dimensional subspace of $\mathfrak{g}$).

**Proposition 4.2.** *The exponential map is a local diffeomorphism around $0 \in \mathfrak{g}$.*

*Proof.* Its differential, if well defined, is the identity on $\mathfrak{g} = T_eG$, so we only need to show that it is a smooth map. The flow on $G \times \mathfrak{g}$ associated to the smooth vector field $(g, X) \mapsto (X(g), 0)$ is given by: $\mathbb{R} \times G \times \mathfrak{g} \ni (t, g, X) \mapsto (g \cdot \exp(tX), X)$, and the flow of a smooth vector field is smooth. Therefore, the exponential map is smooth. □

## 13.5 Proof of Cartan's theorem

Recall the formulation of the theorem:

**Theorem** (Cartan)**.** *Every closed subgroup of a Lie group is a smooth manifold, hence a Lie subgroup.*

*Proof.* Let $H \subset G$ be a closed subgroup of a Lie group. Let $\mathfrak{g}$ denote the Lie algebra of $G$, i.e. the tangent space at the identity. We will define a subspace of $\mathfrak{g}$ which will be the candidate for the tangent space of the identity for $H$. Then we will show that it is indeed so.

Choose a Euclidean metric on $\mathfrak{g}$ and let $\exp : \mathfrak{g} \to G$ be the exponential map. In a neighborhood of the idenity in $\mathfrak{g}$, it is a diffeomorphism onto a neighborhood of the identity in $G$, and let log denote its inverse in that neighborhood.

Let $W \subset \mathfrak{g}$ be the set of all $tX$, where $t \in \mathbb{R}$ and $X \in \mathfrak{g}$ is the limit of a sequence: $\frac{h_n}{|h_n|}$ with $h_n \to 0 \in \mathfrak{g}$ and $\exp(h_n) \in H$. We claim:

1. $\exp(W) \subset H$;
2. $W$ is a linear subspace of $\mathfrak{g}$.

For the first, if $\frac{h_n}{|h_n|} \to X$ and $|h_n| \to 0$ we can choose, for given $t \in \mathbb{R}$, integers $m_n \in \mathbb{Z}$ such that $m_n|h_n| \to t$, so $\exp(m_n \cdot h_n) \to \exp(tX)$ as $n \to \infty$.

Here we will use the following fact: for an one-dimensional subspace of $\mathfrak{g}$ the exponential map is a homomorphism of groups. Therefore, $\exp(m_n \cdot h_n) = \exp(h_n)^{m_n}$, therefore it belongs to $H$. Since $H$ is closed, the limit $\exp(tX)$ is also in $H$.

For the second claim, if $X, Y \in W$ set $h(t) := \log(\exp(tX)\exp(tY))$. We claim that $\lim_{t\to 0} h(t)/t = X + Y$. Indeed, the differential at the identity of the multiplication map: $G \times G \to G$ is $\mathfrak{g} \times \mathfrak{g} \ni (X, Y) \mapsto X + Y$. Hence, $h(t)/|h(t)| = h(t)/t \cdot t/|h(t)| \to \frac{X+Y}{|X+Y|}$ as $t \to 0$, $t > 0$, therefore $X + Y \in W$.

Having proven the two claims, and given that the exponential map is a diffeomorphism in a neighborhood of the identity, it now suffices to show that $\exp(W)$ is a neighborhood of the identity in $H$. Let $D$ be the orthogonal complement of $W$ in $\mathfrak{g}$ with respect to the above norm. For a sequence $h_n \in H$ with $h_n \to e$, we can eventually write $h_n = \exp(x_n + y_n)$ with $x_n \in W$ and $y_n \in D$, $(x_n, y_n) \to 0$. We claim that:

$$\lim_{n\to\infty} \frac{\log(h_n \exp(-x_n))}{|y_n|} = \lim_{n\to\infty} \frac{y_n}{|y_n|}$$

if one of the two limits exists. This is the consequence of the *Campbell-Hausdorff formula*, which expresses the quotient of $\exp(x)\exp(y)$ by $\exp(x + y)$ in terms of commutators (look it up online). The point is that if $x_n$ and $y_n \to 0$, their commutator will go to zero even faster, and this establishes the claim.

□

## 13.6 Morphisms of groups and morphisms of Lie algebras

We have an obvious functor from finite dimensional, differentiable representations of a Lie group $G$ to representations of its Lie algebra, simply by taking

the differential at the identity of the map:

$$G \to \mathrm{GL}(V).$$

Recall that we have shown every finite dimensional representation of a *compact* Lie group to be algebraic, in particular differentiable.

More generally, for every morphism of Lie groups: $G_1 \to G_2$ we get a morphism of Lie algebras: $\mathfrak{g}_1 \to \mathfrak{g}_2$. Vice versa:

**Proposition 6.1.** *Let $G_1, G_2$ be Lie groups with $G_1$ connected and simply connected, then every homomorphism: $\mathfrak{g}_1 \to \mathfrak{g}_2$ lifts to a unique homomorphism: $G_1 \to G_2$.*

This follows from:

**Proposition 6.2.** *Given a Lie group $G$ and a sub-Lie algebra $\mathfrak{h} \subset \mathfrak{g}$, there is a unique connected* immersed Lie subgroup $H \subset G$ *whose Lie algebra is $\mathfrak{h}$.*

By an immersed Lie subgroup we mean an immersed submanifold: $H \to G$ such that $H$ is a subgroup of $G$.

*Proof that Proposition* **??** *implies Proposition* **??**. Given a homomorphism: $\mathfrak{g}_1 \to \mathfrak{g}_2$, compose it with the identity to get $\rho : \mathfrak{g}_1 \to \mathfrak{g} := \mathfrak{g}_1 \oplus \mathfrak{g}_2$.

Proposition **??** gives a unique connected immersed Lie subgroup: $H \to G_1 \times G_2$ whose Lie algebra is $\mathfrak{h}$. Composing with projection to $G_1$ we get: $H \to G_1$ which is an isomorphism on tangent spaces, hence a covering map. Since $G_1$ is simply connected, $H = G_1$. Composing with projection to $G_2$ we get the desired map. □

*Proof of Proposition* **??**. The left translations of $\mathfrak{h}$ give rise to a *distribution* $D_{\mathfrak{h}}$, i.e. a subbundle of $TG$. It is known from the theory of differential equations that a distribution $D$ is (uniquely) *integrable* if and only if for any two vector fields which lie in it, their commutator also lies in it. This is easily seen to be the case for $D_{\mathfrak{h}}$, since $\mathfrak{h}$ is a Lie subalgebra. The leaf through zero of the corresponding foliation is the desired immersed subgroup. □

**Corollary 6.3.** *There is an equivalence of categories between connected, simply connected Lie groups and finite-dimensional Lie algebras.*

*Proof.* The only element of the proof that we are missing is the fact that every finite dimensional Lie algebra is the Lie algebra of a group, which by Proposition **??** can be inferred from the fact that every finite dimensional Lie algebra has a faithful, finite dimensional representation. This uses structure theory of Lie algebras, and we won't prove it. □

# Chapter 14

# Finite-dimensional representations of $\mathfrak{sl}_2(\mathbb{C})$ and of general semisimple Lie algebras

This Lecture is to be read in two parts: the parts on $\mathfrak{sl}_2$ should be read now. Then one builds up the structure theory of general Lie algebras (using the representation theory of $\mathfrak{sl}_2$), to be discussed in the following lectures. Then one discusses general semisimple Lie algebras. However, given the structure theory, the arguments in the general case are so similar to the case of $\mathfrak{sl}_2$ that it is more convenient to place them here.

## 14.1 The Lie algebra $\mathfrak{sl}_2(\mathbb{C})$, and a central element

The Lie algebra of $\mathfrak{sl}_2$ can be identified with the algebra of $2 \times 2$ matrices of trace zero, with Lie bracket the commutator of two matrices. It is generated over the underlying field by three elements $H, E, F$ with bracket relations:

$$[H, E] = 2E,$$

$$[H, F] = -2F,$$

$$[E, F] = H.$$

It is easily verified that the center $\mathcal{Z}(\mathfrak{g})$ of the universal enveloping algebra contains the element:

$$\Delta = 4FE + (H + 2)H.$$

It turns out (but we won't use it – see the Harish-Chandra isomorphism in later lecture) that $\mathcal{Z}(\mathfrak{g})$ is a polynomial ring generated by this element.

In this lecture, all vector spaces are finite dimensional.

## 14.2 Highest weight vectors

Given a representation $V$ of $\mathfrak{sl}_2$, and $\lambda \in \mathbb{C}$, let $V_\lambda$ denote the $\lambda$-eigenspace of $H$. We don't know yet that $H$ acts semisimply, so a priori $V$ is not the direct sum of the $V_\lambda$'s.

**Lemma 2.1.** *$E \cdot V_\lambda \subset V_{\lambda+2}$; $F \cdot V_\lambda \subset V_{\lambda-2}$.*

**Corollary 2.2.** *There is a non-zero vector $v \in V$ which is an eigenvector for $H$ and such that $Ev = 0$.*

This follows by the fact that $V$ is finite dimensional, and the $V_\lambda$'s are linearly independent.

We call $v$ a *highest weight vector*. Similarly, there will be a lowest weight vector, i.e. a nonzero eigenvector of $H$ which is annihilated by $F$.

**Proposition 2.3.** *Fix a heighest weight vector $v \in V_\lambda$, and let $V'$ be the span of $\{F^i v\}_{i\in\mathbb{N}}$. Then $V'$ is $\mathfrak{sl}_2$-stable, irreducible, and $\Delta$ acts by $\lambda(\lambda_2)$. The highest weight $\lambda$ is a non-negative integer, and $V'$ is the sum of one-dimensional weight spaces $V'_\mu$ for $\mu = \lambda, \lambda-2, \lambda-4, \ldots, -\lambda$.*

*Proof.* It is clearly stable under $F$ and $H$. We easily compute:

$$EF^n v_\lambda = n(\lambda - (n-1))F^{n-1}v_\lambda.$$

Hence, the space is $E$-stable.

Moreover, since it is finite-dimensional, we must have $n(\lambda-(n-1)) = 0$ for some $n \geqslant 1$, hence $\lambda$ is a non-negative integer. In that case, $n = \lambda + 1$, and $F^n v_\lambda$ must be zero (because it is a highest weight vector of weight $-\lambda - 2$ and, by the same argument, it cannot generate a finite-dimensional representation). On the other hand, for $n < \lambda + 1$ $EF^n v_\lambda \neq 0$, hence $F^n v_\lambda \neq 0$. The statement about the weight spaces of $V'$ follows.

We have: $\Delta v = 4FEv + (H+2)Hv = 0 + \lambda(\lambda+2)v$. Since $\Delta$ commutes with the action of $\mathfrak{sl}_2$ and is generated by $v$, all elements of $V'$ have the same $\Delta$-eigenvalue.

On the other hand, $V'$ has at most one eigenvector for each $H$-eigenvalue. If $V'$ was reducible, there would be some highest weight vector with eigenvalue $\neq$ □

**Corollary 2.4.** *Irreducible (finite-dimensional) representations of $\mathfrak{sl}_2$ are $H$-semisimple.*

We will eventually see that all finite-dimensional representations of $\mathfrak{sl}_2$ are semisimple, in particular $H$-semisimple.

**Lemma 2.5.** *For every nonnegative integer $n$ there is an irreducible finite-dimensional representation of heighest weight $n$. It is unique up to isomorphism and has dimension $n+1$.*

*Proof.* If $V$ denotes the standard, 2-dimensional representation, then it is easy to see that $S^nV$ has a unique highest weight vector with weight $n+1$, hence is irreducible. Uniqueness follows from the explicit description of the action of $E, F$ and $H$ above. □

This existence statement will require a lot more work in the general case.

## 14.3 Semisimplicity (complete reducibility)

Suppose that we have a short exact sequence of representations: $0 \to A \to B \to C \to 0$. How do we know that it splits? Well, it splits as vector spaces. That is, there is an element of $\mathrm{Hom}_{\mathbb{C}}(C, B)$ which lifts the identity element in $\mathrm{Hom}_{\mathbb{C}}(C, C)$. We would like to know that there is a $\mathfrak{g}$-invariant such element. Thus, it suffices to show that if we apply the functor of "$\mathfrak{g}$-invariants" to the exact sequence:

$$0 \to \mathrm{Hom}_{\mathbb{C}}(C, A) \to \mathrm{Hom}_{\mathbb{C}}(C, B) \to \mathrm{Hom}_{\mathbb{C}}(C, C) \to 0,$$

it remains exact.

This is a problem is cohomology. Any short exact sequence of $\mathfrak{g}$-modules $0 \to U \to V \to W \to 0$ (think of the above Hom spaces here) gives rise to a long exact sequence:

$$0 \to U^{\mathfrak{g}} \to V^{\mathfrak{g}} \to W^{\mathfrak{g}} \to H^1(\mathfrak{g}, U) \to H^1(\mathfrak{g}, V) \to H^1(\mathfrak{g}, W) \to \dots.$$

The right derived functor $H^1$ can be explicitly described, it turns out, as follows (more details on Lie algebra cohomology, hopefully, in some future version of these notes):

$$H^1(\mathfrak{g}, V) = Z^1(\mathfrak{g}, V)/C^1(\mathfrak{g}, V),$$

where the cocycles $Z^1(\mathfrak{g}, V)$ are maps $f : \mathfrak{g} \to V$ satisfying:

$$f([X, Y]) = Xf(Y) - Yf(X),$$

and the coboundaries are those of the form: $f(X) = Xv$ (for some $v \in V$).

**Theorem 3.1.** *For any finite-dimensional $\mathfrak{g}$-module $V$, $H^1(\mathfrak{g}, V) = 0$.*

*Proof.* First, we reduce to simple $\mathfrak{g}$-modules by induction. Suppose that we have a short exact sequence:

$$0 \to U \to V \to W \to 0,$$

and that the first cohomology groups of $U$ and $W$ are trivial, then the long exact sequence shows that $H^1(\mathfrak{g}, V) = 0$, as well.

For a simple $V$, if $V = \mathbb{C}$ then $H^1(\mathfrak{g}, V) = 0$ trivially.

If $V \neq \mathbb{C}$ then, as we saw in $\Delta$ acts by a non-zero scalar. Given a cocycle $f$, let $v \in V$ be defined by the equation:

$$\frac{1}{8}\Delta v = \sum_i X_i f(Y_i),$$

where $(X_i)_i, (Y_i)_i$ are dual bases for $\mathfrak{g}$ under the Killing form. We will use the fact that $\Delta = 8(\sum_i X_i Y_i)$, in the case of $\mathfrak{sl}_2$, and in the general case this sum is called the *Casimir* operator, does not depend on the choice of basis and lies in the center of $\mathfrak{z}(\mathfrak{g})$.

One can now easily show that $\Delta f(X) = \Delta X v = X \Delta v$, which implies that $f(X) = Xv$, for all $X \in \mathfrak{g}$. □

**Corollary 3.2.** *Every finite-dimensional $\mathfrak{g}$-module is semisimple.*

## 14.4 General $\mathfrak{g}$

There is very little that changes in the case of a general $\mathfrak{g}$. Choosing a Borel subalgebra $\mathfrak{b}$ and a Cartan subalgebra $\mathfrak{h}$ thereof, denoting by $(\mathfrak{sl}_2)_\alpha$ the subalgebras $< H_\alpha > + \mathfrak{g}_\alpha + \mathfrak{g}_{-\alpha}$, for every positive root $\alpha$, and by $\mathfrak{n}_-$, $\mathfrak{n}_+$ the sums of positive/negative weight spaces, we can easily prove as before (namely, using the fact that each nonzero element of $\mathfrak{n}_+$ raises the weight):

**Lemma 4.1.** *Every finite dimensional representation $V$ contains a heighest weight vector, i.e. an eigenvector for $\mathfrak{b}$.*

Obviously, irreducible representations contain a unique highest-weight vector (because the $\mathfrak{g}$-span of a highest weight vector contains vectors of smaller weight only).

Applying Proposition **??** to the action of $(\mathfrak{sl}_2)_\alpha$ we get:

**Proposition 4.2.** *If $\lambda \in \mathfrak{h}^*$ is the weight of a highest weight vector then $\langle \lambda, \check{\alpha} \rangle \in \mathbf{N}$ for every root $\alpha > 0$. The action of $\mathfrak{h}$ on a finite-dimensional representation $V$ is semisimple, and for every element $w \in W$ the weight spaces $V_\mu$ and $V_{w\mu}$ have the same dimension.*

*Proof.* The first statement follows directly from Proposition **??**.

For the second, let $V'$ the span of $\mathfrak{h}$-semisimple vectors in the $\mathfrak{g}$-span of a highest weight vector. I claim that $V'$ is $\mathfrak{g}$-stable. Indeed, we have a homomorphism of $\mathfrak{h}$-modules: $\mathfrak{g} \otimes V' \to V$, and the left hand side is $\mathfrak{h}$-semisimple. Therefore, its image belongs to $V'$.

Finally, the last statement is enough to prove for simple reflections, and in that case it follows from the analogous statement for $(\mathfrak{sl}_2)_\alpha$-representations. □

We now discuss the Casimir operator $C = \sum_i X_i Y_i$, where $(X_i)_i, (Y_i)_i$ are dual bases with respect to the Killing form. First of all:

**Lemma 4.3.** *$C$ does not depend on the choice of basis. It belongs to the center of $U(\mathfrak{g})$.*

*Proof.* A better way to define $C$ is as follows: the Killing form is an invariant 2-tensor on $\mathfrak{g}^*$, i.e. an element of:

$$\left(\otimes^2 \mathfrak{g}^*\right)^{\mathfrak{g}}.$$

It is nondegenerate and invariant, hence induces an isomorphism of $\mathfrak{g}$-modules: $\mathfrak{g}* \to \mathfrak{g}$. Its image under:

$$\otimes^\bullet \mathfrak{g}^* \to \otimes^\bullet \mathfrak{g} \to U(\mathfrak{g})$$

is the Casimir operator. This proves the lemma. □

Now:

**Lemma 4.4.** *The Casimir operator acts on an irreducible representation of highest weight $\lambda$ by the scalar:*

$$(\lambda + \rho, \lambda + \rho) - (\rho, \rho),$$

*where $(\ ,\ )$ denotes the Killing form and $\rho$ is half the sum of positive roots.*

*Proof.* The trick is to write the operator as an element of $U(\mathfrak{h})$ plus an element in the ideal generated by $\mathfrak{n}^+$. The latter will kill $v_\lambda$ (the highest weight vector), and the former will give the eigenvalue.

We have an orthogonal decomposition: $\mathfrak{g} = \mathfrak{h} \oplus \sum_{\alpha>0}(\mathfrak{g}_\alpha \oplus \mathfrak{g}_{-\alpha})$; if $X_\alpha \in \mathfrak{g}_\alpha, Y_\alpha \in \mathfrak{g}_{-\alpha}$ are dual elements, the Casimir element will be equal to an element $Z\ in U(\mathfrak{h})$ plus:

$$\sum_{\alpha>0} (X_\alpha Y_\alpha + Y_\alpha X_\alpha) = \sum_{\alpha>0} (2Y_\alpha X_\alpha + [X_\alpha, Y_\alpha]).$$

If we apply this to $v_\lambda$ we will get $\lambda\left(\sum_{\alpha>0}[X_\alpha, Y_\alpha]\right) v_\lambda$.

Notice that for $H \in \mathfrak{h}$, $(H, [X_\alpha, Y_\alpha]) = ([H, X_\alpha], Y_\alpha) = \alpha(H)(X_\alpha, Y_\alpha) = \alpha(H)$, and therefore $[X_\alpha, Y_\alpha]$ is the image of $\alpha$ under the identification: $\mathfrak{h}^* \to \mathfrak{h}$ induced by the Killing form. Therefore,

$$\lambda\left(\sum_{\alpha>0}[X_\alpha, Y_\alpha]\right) = \sum_{\alpha>0}(\lambda, \alpha) = 2(\lambda, \rho).$$

On the other hand, the element $Z \in U(\mathfrak{h})$ is the restriction of the Killing form to $\mathfrak{h}$, interpreted as in the proof of Lemma **??**: via the map $\otimes^\bullet \mathfrak{h}^* \to \otimes^\bullet \mathfrak{h} \to U(\mathfrak{h})$. This means that its evaluation on $\lambda$ is the Killing form $(\lambda, \lambda)$.

Hence we get that the Casimir acts on $v_\lambda$ by the scalar: $(\lambda, \lambda) + 2(\lambda, \rho) = (\lambda + \rho, \lambda + \rho) - (\rho, \rho)$. □

Finally, vanishing of cohomology and complete reducibility are proven as previously, i.e. Theorem **??** and Corollary **??** hold for all semisimple $\mathfrak{g}$.

*Remark.* All the above could be done with an arbitrary nondegenerate invariant bilinear form, not necessarily the Killing form, replacing the Casimir element accordingly. But on each simple factor, such a form is necessarily a multiple of the Killing form. Moreover, Theorem **??** can be proven using the invariant form $(X, Y) \mapsto \mathrm{tr}_V(X, Y)$ which vanishes on the kernel of the representation $V$.

# Chapter 15

# Structure of general (finite dimensional) Lie algebras

Later

# Chapter 16

# Structure of semisimple Lie algebras

In fact, the present lecture also contains some general theorems for the existence and conjugacy of Cartan subalgebras and Borel subalgebras, whose proof, however, is eventually reduced to the semisimple case.

## 16.1 Jordan decomposition in $\mathfrak{gl}$.

**Definition.** Let $\mathfrak{g}$ be a Lie algebra. An element $X \in \mathfrak{g}$ is called *semisimple* if $\mathrm{ad}(X)$ is a semisimple operator, and *nilpotent* if $\mathrm{ad}(X)$ is nilpotent.

*Remark.* For the Lie algebra $\mathfrak{g} = \mathrm{End}(V)$, where $V$ is a vector space, these notions of semisimple and adjoint do not completely coincide with "semisimple operator" and "nilpotent operator", but they coincide modulo the center of $\mathfrak{g}$, as the following result shows. For this case, we will be using the word "semisimple" or "adjoint" to refer to the property of the operator, unless otherwise stated.

**Proposition 1.1.** *An operator $T \in \mathrm{End}(V)$ is semisimple iff the operator $\mathrm{ad}(T) \in \mathrm{End}(\mathrm{End}(V))$ is semisimple. It is nilpotent in $\mathrm{End}(V)/k$ (where $k$ stands for the center of $\mathrm{End}(V)$ iff $\mathrm{ad}(T)$ is nilpotent.*

Now recall that for any element $X \in \mathrm{End}(V)$ (some vector space $V$) there is a unique *Jordan decomposition* $X = X_s + X_n$ with $X_s$ semisimple, $X_n$ nilpotent and $[X_s, X_n] = 0$. (Moreover, $X_s$ and $X_n$ commute with every element commuting with $X$, since they can be expressed as polynomials in $X$.) Moreover, there are polynomials $P, Q \in k^{p^{-\infty}}[T]$ (where $k$ is the base field, $p$ the characteristic) such that $X_s = P(X)$ and $X_n = Q(X)$ – in particular, *$X_s$ and $X_n$ are defined over $k^{p^{-\infty}}$*. (Indeed, if the characteristic polynomial over the algebraic closure is written $\prod_i (T - a_i)_i^m$, with all $a_i$ distinct, choose $P(T)$ to satisfy the congruences:

$$P(T) \equiv a_i \mod (T - a_i)_i^m,\ P(T) \equiv 0 \mod T.)$$

## 16.2 Derivations and the Jordan decomposition

**Definition.** A *derivation* of a Lie algebra $\mathfrak{g}$ is a linear map $D : \mathfrak{g} \to \mathfrak{g}$ satisfying $D([X, Y]) = [X, D(Y)] + [D(X), Y]$.

*Remarks.* 1. This is a very natural extension of the definition of derivation for an associative algebra, since such a derivation induces a derivation as above on the associated Lie algebra. Vice versa, a derivation of a Lie algebra induces a derivation of its universal enveloping algebra.

2. Derivations form a Lie subalgebra of $\mathrm{End}(\mathfrak{g})$.

3. The adjoint representation $\mathrm{ad} : \mathfrak{g} \to \mathrm{End}(\mathfrak{g})$ has image in $\mathrm{Der}(\mathfrak{g})$.

**Proposition 2.1.** *Every derivation of a* semisimple *Lie algebra is* inner*, i.e. in the image of* ad*.*

*Proof.* The formula $[D, \mathrm{ad}(X)] = \mathrm{ad}(DX)$ shows that the image of ad is an ideal in $\mathrm{Der}(\mathfrak{g})$. Since the image is a semisimple Lie algebra, there is a complementary ideal $I$ (namely, its orthogonal complement under the Killing form on $\mathrm{Der}(\mathfrak{g})$). But if $D \in I$, and $I$ is an ideal, the same formula shows that $\mathrm{ad}(DX) \in I \cap \mathrm{ad}(\mathfrak{g}) = 0$, which since ad is injective means that $DX = 0$, i.e. $D = 0$. □

**Corollary 2.2.** *The identity component of the automorphism group of a semisimple Lie group coincides with the group of inner automorphisms.*

**Proposition 2.3.** *If* $D \in \mathrm{Der}(\mathfrak{g})$ *then* $D_s, D_n \in \mathrm{Der}(\mathfrak{g})$.

*Proof.* If $X$ is in the generalized $\lambda$-eigenspace and $Y$ is in the generalized $\mu$-eigenspace for $D$, then it can be shown by induction that:

$$(D - (\lambda + \mu))^n([X, Y]) = \sum_{r=0}^{n} \binom{n}{r} [(D - \lambda)^r(X), (D - \mu)^{n-r}(Y)],$$

hence $[X, Y]$ is in the generalized $\mu + \lambda$-eigenspace. This shows that $D_s$ is a derivation, and then $D_n = D - D_s$ is a derivation. □

**Theorem 2.4.** *Let* $\mathfrak{g}$ *be a semisimple Lie algebra. Then every element has a unique decomposition (over* $k^{p^{-\infty}}$*),* $X = X_s + X_n$ *with* $X_s$ *semisimple,* $X_n$ *nilpotent and* $[X_s, X_n] = 0$*.*

*Proof.* By the previous two propositions, $\mathrm{ad}(X)_s$ and $\mathrm{ad}(X)_n$ are derivations and therefore belong to the image of ad. This proves the existence (and uniqueness) of $X_s$ and $X_n$. □

If we assume complete reducibility of finite-dimensional representations of semisimple Lie algebras (which we will prove later), we can show that the Jordan decomposition is preserved by homomorphisms of semisimple Lie algebras. In fact, we can show more generally:

**Theorem 2.5.** *For a homomorphism $\rho : \mathfrak{g} \to \mathrm{End}(V)$ we have $\rho(X_s) = \rho(X)_s$, $\rho(X_n) = \rho(X)_n$.*

*Proof.* By complete reducibility of $\mathrm{End}(V)$ under the adjoint $\mathfrak{g}$-action, we have:

$$\mathrm{End}(V) = \rho(\mathfrak{g}) \oplus \mathfrak{m},$$

where $\mathfrak{m}$ is an $\mathrm{ad}(\rho(\mathfrak{g}))$-invariant subspace. (Notice that in Proposition **??** we were able to obtain a similar decomposition in $\mathrm{Der}(\mathfrak{g})$ by using the Killing form, so we did not need to know reducibility.)

Since $\rho(X)_s, \rho(X)_n$ are polynomials in $\rho(X)$, their adjoint action preserves both $\rho(\mathfrak{g})$ and $\mathfrak{m}$. Let $\rho(X)_n = \rho(a) + b$ with $a \in \mathfrak{g}, b \in \mathfrak{b}$. Then $[\rho(\mathfrak{g}), b] = 0$, which means that $b \in \mathrm{End}(V)$ is a $\mathfrak{g}$-endomorphism. If $V = \oplus V_i$ is a decomposition into irreducibles, $b$ acts by a scalar on each one of them, by Schur's lemma. On the other hand, we know that $\rho(X)_n$ is nilpotent, $\rho(a)$ and $b$ commute, and $\mathrm{tr}_{V_i}(\rho(a)) = 0$ because $a$ (like every element of $\mathfrak{g}$) is a sum of commutators. Therefore, $\mathrm{tr}_{V_i}(b) = 0$, hence $b$ acts by zero on all $V_i$, i.e. $b = 0$.

Now, $\rho(X)_n = \rho(a)$ acts nilpotently on $V$, hence it acts nilpotently on $\mathrm{End}(V)$ under the adjoint representation. By the decomposition $\mathrm{End}(V) = \mathfrak{g} \oplus \mathfrak{m}$ it follows that it acts nilpotently on $\mathfrak{g}$. By the uniqueness of the Jordan decomposition we can now infer that $\rho(X)_n = \rho(X_n)$. □

We will see later that the image of a semisimple element of $\mathfrak{g}$ (where $\mathfrak{g}$ is semisimple) under any representation is a semisimple operator, and the image of a nilpotent element is a nilpotent operator.

## 16.3 Cartan subalgebras

A *Cartan subalgebra* of a Lie algebra $\mathfrak{g}$ is a *nilpotent, self-normalizing* subalgebra $\mathfrak{h}$. Here $\mathfrak{g}$ is not (yet) necessarily semisimple.

We will construct Cartan subalgebras as nilspaces (generalized eigenspaces of zero under the adjoint representation) of *s-regular* elements. Then we will show that they are all conjugate to each other.

**Definition.** An s-regular element $X \in \mathfrak{g}$ is an element with minimal possible 0-generalized eigenspace under ad.

Since the dimension of the zero generalized eigenspace is the heighest power of $t$ which divides the characteristic polynomial of $\mathrm{ad}(X)$, it follows that s-regular elements form a (nonempty) Zariski open set.

*Remark.* In many books on Lie algebras, the word "regular" is used for "s-regular" (which is my invention!). The problem with this is that it has become nowadays standard to call "regular" the elements with minimal *zero eigenspace*, i.e. *centralizer*, instead of generalized eigenspace. This includes non-semisimple elements, while s-regular, as we shall see, implies semisimple (for semisimple Lie algebras) – hence for those: s-regular = regular semisimple.

**Proposition 3.1.** *The nilspace of a s-regular element is a Cartan subalgebra.*

*Proof.* Let $X$ be the s-regular element and $\mathfrak{h}$ its centralizer. We will prove that $\mathfrak{h}$ is nilpotent; equivalently, by Engel's theorem, that the restriction of $\mathrm{ad}(Y)$ to $\mathfrak{h}$, for any $Y \in \mathfrak{h}$, is nilpotent. Let $U \subset \mathfrak{h}$ be the subset of elements which fail to satisfy this; it is a Zariski open subset (again by considerations of the characteristic polynomial). Let $V \subset \mathfrak{h}$ be the subset of elements which act invertibly on $\mathfrak{g}/\mathfrak{h}$. It is again a Zariski open subset, and non-empty since $X \in V$. If $U \neq \varnothing$ then $U \cap V \neq \varnothing$, i.e. there exists an element $Y \in \langle$ such that the dimension of the zero generalized eigenspace for $Y$ is less than the dimension of $\mathfrak{h}$, a contradiction by the s-regularity of $X$. Thus, $\mathfrak{h}$ is nilpotent.

If $Z$ normalizes $\mathfrak{h}$ then $[Z, X] \in \mathfrak{h}$ which implies that $Z$ is in the nilspace of $X$, i.e. in $\mathfrak{h}$. □

Hence, every Lie algebra has Cartan subalgebras. We will eventually prove that any two Cartan subalgebras are conjugate (over the algebraic closure) by the group of inner automorphisms of $\mathfrak{g}$ and, in particular, equal to nilspaces of s-regular elements.

## 16.4 Root decomposition, semisimple case

**Theorem 4.1.** *Assume that $\mathfrak{g}$ is semisimple, and let $\mathfrak{h}$ be the nilspace of an s-regular element (hence[1] a Cartan subalgebra). Then:*

1. *$\mathfrak{h}$ is abelian.*
2. *The centralizer of $\mathfrak{h}$ is $\mathfrak{h}$.*
3. *Every element of $\mathfrak{h}$ is semisimple.*
4. *The restriction of the Killing form (or any non-degenerate invariant form) of $\mathfrak{g}$ to $\mathfrak{h}$ is non-degenerate.*

*Proof.* The rest of the statements follow from the last one. Let us see how:

Cartan's criterion says that a Lie subalgebra $\mathfrak{a}$ of $\mathrm{End}(V)$ is solvable if and only if $\mathrm{tr}(XY) = 0$ for every $X \in \mathfrak{a}, Y \in [\mathfrak{a}, \mathfrak{a}]$. Applying this to $\mathrm{ad}(\mathfrak{h}) \subset \mathrm{End}(\mathfrak{g})$ (which is nilpotent, hence solvable), we get that $B(X, Y) = 0$ for all $X \in \mathfrak{h}, Y \in [\mathfrak{h}, \mathfrak{h}]$ (where $B$ is the Killing form for $\mathfrak{g}$). Therefore, the radical of the restriction of $B$ to $\mathfrak{h}$ contains the commutator, which means that $[\mathfrak{h}, \mathfrak{h}] = 0$.

The centralizer is contained in the normalizer, which is $\mathfrak{h}$, but since $\mathfrak{h}$ is abelian it coincides with it.

Finally, let $X \in \mathfrak{h}$ and let $X = X_s + X_n$ be its Jordan decomposition. Since $X_s, X_n$ commute with the centralizer of $X$, which contains $\mathfrak{h}$, it follows that $X_s, X_n$ are in the centralizer of $\mathfrak{h}$, which is $\mathfrak{h}$. Thus, if $Y \in \mathfrak{h}$, $\mathrm{ad}(Y)\mathrm{ad}(X_n)$ is nilpotent, which implies that $\mathrm{ad}(X_n)$ is orthogonal to $\mathfrak{h}$ under the Killing form. By non-degeneracy of the Killing form on $\mathfrak{h}$, $X_n = 0$.

[1] Eventually, since they are conjugate, all Cartan subalgebras are of this form

We come to the proof of the last statement: if $X$ is a regular element such that $\mathfrak{g}$ is (the nilspace of $X$), let $\mathfrak{g} = \bigoplus_\lambda \mathfrak{g}_\lambda$ be a decomposition of $\mathfrak{g}$ into generalized $\mathrm{ad}(X)$-eigenspaces. As we saw in the proof of Proposition **??**, $[\mathfrak{g}_\lambda, \mathfrak{g}_\mu] \subset \mathfrak{g}_{\lambda+\mu}$, which implies that $\mathfrak{g}_\lambda \perp \mathfrak{g}_\mu$ (under the Killing form), unless $\lambda + \mu = 0$. Therefore, the decomposition:

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \bigoplus (\mathfrak{g}_\lambda \oplus \mathfrak{g}_{-\lambda})$$

is orthogonal, and since $B$ is nondegenerate, it has to be non-degenerate on each of the summands, in particular on $\mathfrak{h} = \mathfrak{g}_0$. □

**Corollary 4.2.** *$\mathfrak{h}$ is a maximal abelian subalgebra of $\mathfrak{g}$. Every s-regular element is semisimple.*

The decomposition of $\mathfrak{g}$ into generalized eigenspaces for the adjoint action of a Cartan subalgebra $\mathfrak{h}$ is called the *Cartan decomposition* of $\mathfrak{g}$:

$$\mathfrak{g} = \mathfrak{h} + \sum_{\alpha \Phi \subset \mathfrak{h}^*} \mathfrak{g}_\alpha.$$

The set $\Phi$ of non-zero elements of $\mathfrak{h}^*$ which appear in this decomposition is called the set of *roots* of $\mathfrak{g}$.

Now I collect quickly the basic facts about the root decomposition, which are well-known and easy to follow in the literature. Only hints about the proofs are given.

- $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subset \mathfrak{g}_{\alpha+\beta}$.
- $\Phi$ spans $\mathfrak{h}^*$.
- If $\alpha \in \Phi$, let $\langle_\alpha \in \mathfrak{h}$ be the image of $\alpha$ under the isomorphism: $\mathfrak{h}^* \to \mathfrak{h}$ defined by a non-degenerate invariant form $(\ ,\ )$ on $\mathfrak{g}$. Then, for all $H \in \mathfrak{h}$, $X \in \mathfrak{g}_\alpha, Y \in \mathfrak{g}_{-\alpha}$ we have: $(H, [X,Y]) = ([H,X], Y) = \alpha(H)(X,Y) = (H, (X,Y)h_\alpha)$, and therefore: $[X,Y] = (X,Y)h_\alpha$. Notice that the pairing $(\ ,\ )$ between $\mathfrak{g}_\alpha$ and $\mathfrak{g}_{-\alpha}$ is nonzero because otherwise it would be degenerate on $\mathfrak{g}$.

  It follows that $\mathfrak{h}_\alpha := [\mathfrak{g}_\alpha, \mathfrak{g}_{-\alpha}]$ is one-dimensional, spanned by the element $h_\alpha$. We let $H_\alpha$ (or $\check{\alpha}$) denote its unique multiple with $\alpha(H_\alpha) = 2$ (the *coroot* associated to $\alpha$).
- The sum $\mathfrak{h}_\alpha + \mathfrak{g}_\alpha + \mathfrak{g}_{-\alpha}$ is a subalgebra isomorphic to $\mathfrak{sl}_2$. To prove this, one first shows that it includes an $\mathfrak{sl}_2$-triple $(H, E, F)$. An element of $\mathfrak{g}_{-\alpha}$ orthogonal to $E$ would be a highest weight vector of weight $-2$, which is absurd. Therefore, $\mathfrak{g}_{-\alpha}$ is one-dimensional, and so is $\mathfrak{g}_\alpha$.
- If $\alpha, c\alpha \in \Phi$ then $c = \pm 1$.
- For $\alpha \in \Phi$, the reflection $\mathfrak{h}^* \ni \gamma \mapsto \gamma - \langle \gamma, H_\alpha \rangle\, \alpha$ fixes $\Phi$.

  Indeed, this follows from viewing $\mathfrak{g}$ as an $\mathfrak{sl}_2$-module: a nonzero element $Y$ of $\mathfrak{g}_\gamma$ has weight $\langle \gamma, H_\alpha \rangle$, and by properties of $\mathfrak{sl}_2$-modules the element $F^p Y$ is nonzero. But that lives in $\mathfrak{g}_{\gamma - \langle \gamma, H_\alpha \rangle \alpha}$.

## 16.5 Conjugacy of Borel subalgebras and the universal Cartan: statements

A maximal solvable subalgebra of a Lie algebra (over the algebraic closure) is called a *Borel subalgebra*. Clearly, a Borel subalgebra contains the radical of $\mathfrak{g}$, and therefore most questions (most importantly, the question of conjugacy) are reduced to the semisimple case.

The following is immediate:

**Lemma 5.1.** *Every Borel subalgebra is its own normalizer.*

Let $G$ be the group of *inner automorphisms* of $\mathfrak{g}$, i.e. automorphisms of the form $\exp(\mathrm{ad}(y)) \in \mathrm{GL}(\mathfrak{g})$. Here $\mathfrak{g}$ is not assumed to be semisimple, although clearly the proof of the following theorem reduces to the semisimple case:

**Theorem 5.2.** *All Borel subalgebras are $G$-conjugate.*

*Remark.* The definition of $G$ is not satisfactory, because it relies on the exponential map which is not algebraic. Here is how we would define an algebraic subgroup of $\mathrm{GL}(V)$ in general, in characteristic zero: We would take $G$ to be the subgroup generated by $\exp(\mathrm{ad}(X))$ for all *nilpotent* elements of $X$ (the exponential is algebraic on them!). It turns out, for semisimple groups at least (which is all we care about in this theorem) that these generate the same group as the one analytically defined above. In arbitrary characteristic, over an algebraically closed field $k$, an algebraic group $G(k)$ acts transitively on the Borel subalgebras of its Lie algebra.

The proof will use many lemmas, including:

**Lemma 5.3.** *All Cartan subalgebras of a solvable Lie algebra are conjugate.*

(We write "conjugate" for the group of inner automorphisms of the given Lie algebra, i.e. in the last lemma the inner automorphisms of the solvable algebra. Again, exponentials of nilpotent elements will suffice, although in this case they do not generate the whole group of inner automorphisms.)

**Corollary 5.4.** *All Cartan subalgebras of a Lie algebra $\mathfrak{g}$ are conjugate (over the algebraic closure).*

*Proof of the corollary.* Any Cartan subalgebra is nilpotent, hence solvable, hence contained (over the algebraic closure) in a Borel subalgebra. Two Borel subalgebras are conjugate, and two Cartan subalgebras of a given Borel are conjugate. □

**Definition.** The *universal Cartan* $\mathfrak{h}$ of $\mathfrak{g}$ is the quotient of any Borel subalgebra $\mathfrak{b}$ by its nilpotent radical. It is a commutative Lie algebra. For two different Borel subalgebras $\mathfrak{b}$ and $\mathfrak{b}'$, we identify the corresponding quotients $\mathfrak{h}$ and $\mathfrak{h}'$ by picking an element $g \in G$ which conjugates $\mathfrak{b}$ to $\mathfrak{b}'$; since $\mathfrak{b}$ is its own normalizer, such an identification is unique up to an inner automorphism of $\mathfrak{b}$, but inner automorphisms act trivially on the quotient $\mathfrak{h}$; therefore, $\mathfrak{h}$ *is unique up to unique isomorphism.*

The universal Cartan comes with a set of roots $\Phi \subset \mathfrak{h}^*$, and in fact *with a canonical choice of positive roots* $\Phi^+$ (those that appear in the decomposition of $\mathfrak{b}$). This will play a role in the definition of the Langlands dual group, which has, by definition a canonical maximal torus with Lie algebra isomorphic to $\mathfrak{h}^*$.

## 16.6 The scheme of Borel subgroups

A maximal solvable subgroup of an algebraic group over an algebraically closed field $k$ is called a *Borel subgroup*. For a smooth group scheme $G$ (of finite presentation) over an arbitrary base $S$, we call *Borel subgroup* any smooth subgroup scheme of finite presentation $B$ of $G$ over $S$ such that for all $s \in S$ the geometric fiber $B_{\bar{s}}$ is a Borel subgroup (SGA3, XXIV, 4.5).

In general, it is not true that the functor which assigns to any $S$-scheme $T$ the set of Borel subgroups over $T$ is reprentable, i.e. that there exists a scheme $\mathcal{B}$ over $S$ such that $\mathcal{B}(T)$ is the set of Borel subgroups over $T$. For instance, $G$ would act by automorphisms on such a scheme, and the kernel would be equal to the radical of $G$, but if $S = \operatorname{spec} k$ and $k$ is not perfect then the radical may not be defined over $k$.

However, if $G$ is reductive then it is proven in SGA3, XXVI, 3.3:

**Theorem 6.1.** *If $G$ is a reductive subgroup (i.e. smooth, and all geometric fibers are reductive) over a base $S$, the functor which assigns to an $S$-scheme $T$ the set of Borel subgroups of $G$ over $T$ is representable by a smooth, projective, geometrically integral group scheme $\mathfrak{B}$ over $S$.*

If there is a Borel subgroup $B$ over $S$ then $\mathcal{B} \simeq G/B$ under the action of $G$. The implications of this theorem are very important; for instance:

**Corollary 6.2.** *If $G$ is a reductive group defined over a global field $k$, then it is quasi-split (i.e. has a Borel) over almost all completions $k_v$ of $k$.*

*Proof.* Indeed, for a finite set of places $V$ there is a reductive model of $G$ over the $V$-integers $\mathfrak{o}_V$. Consider the scheme $\mathfrak{B}$ of Borel subgroups over $S = \operatorname{spec} \mathfrak{o}_V$. Let $R$ be a non-archimedean completion of $\mathfrak{o}_V$ and let $F$ be the residue field of $R$ (a finite field). Since $G$ is reductive over $F$, it is known that it has a Borel subgroup defined over $F$. In other words, the scheme $\mathcal{B}$ has an $F$-point. A smooth scheme is *formally smooth*, which in our case implies the henselian property: every $F$-point of $\mathcal{B}$ lifts to an $R$-point. In other words, there is a Borel subgroup at every non-archimedean place outside of $V$. □

## 16.7 Positive roots and standard Borel subgroups

We return to working over an algebraically closed field.

The following are combinatorial properties of root systems:

**Proposition 7.1.** *Let $(V, \Phi, s_\alpha)$ be a root system, let $l$ be a functional which does not vanish on $\Phi$ and let $\Phi^+$ be the elements of $\Phi$ on which $l$ is positive.*

*This is called a choice of* positive roots. *Let $\Delta$ denote the set of elements of $\Phi^+$ which cannot be written as sums of other elements of $\Phi^+$ with positive integral coefficients. (These are the* simple roots *for this choice of positive roots.) Then:*

1. *Every $\alpha \in \Phi^+$ is a sum of elements of $\Delta$.*
2. *If $\alpha, \beta \in \Delta$ then $(\alpha, \beta) \leqslant 0$.*
3. *The elements of $\Delta$ are linearly independent.*

*Proof.* The first property follows immediately from the definition of $\Delta$. The second is proved by classifying root systems of rank 2. For the third, if we had a linear relation $\sum_{\alpha_i \in \Delta} c_i \alpha_i = 0$ then some of the $c_i$'s have to be negative, say for $i = 1, \ldots, r$. Then:

$$\sum_{i=1}^{r} (-c_i)\alpha_i = \sum_{i=r+1}^{|\Delta|} c_i \alpha_i$$

$$\Rightarrow \left\| \sum_{i=1}^{r} (-c_i)\alpha_i \right\|^2 = \left\langle \sum_{i=1}^{r} (-c_i)\alpha_i, \sum_{i=r+1}^{|\Delta|} c_i \alpha_i \right\rangle$$

which by the second statement is less or equal to zero, a contradiction. □

We call *standard Borel subalgebra* a subalgebra of the form $\mathfrak{h} + \sum_{\alpha \in \Phi^+} \mathfrak{g}_\alpha$, where $\mathfrak{h}$ is a Cartan subalgebra and $\Phi^+$ is some choice of positive roots. Of course, in the end it will turn out that every Borel subalgebra is standard.

The proof of conjugacy of Borel subalgebras goes by decreasing induction on the dimension of $\mathfrak{b} \cap \mathfrak{b}'$, where $\mathfrak{b}$ is assumed to be standard, the case where one is included in the other being obvious.

[TO BE CONTINUED]

# Chapter 17

# Verma modules and the category $\mathcal{O}$.

## 17.1 Verma modules

We have seen[1] that finite-dimensional representations of semisimple Lie algebras are completely reducible, and the irreducibles are generated by a highest weight vector (with dominant, integral weight). We now want to *construct* those irreducible representations (in particular, to show that there is a unique one up to unique isomorphism for each given weight), and to compute their *characters*.

In specific cases one can do that "by hand", constructing first the irreducible representations fundamental weights, and then the rest by taking tensor products of those (and subtracing copies of the representations already constructed). For instance, for $\mathfrak{sl}_n$ the $n-1$ fundamental representations are the first $n-1$ exterior powers of the standard, $n$-dimensional represntation.

For a more systematic approach, it is better to move outside the realm of finite-dimensional representations, constructing the universal objects with highest weight.

More precisely, we consider the category of $\mathfrak{g}$-modules of arbitrary, possibly infinite, dimension (no topology), and for $\lambda \in \mathfrak{h}^*$ (where $\mathfrak{h}$ denotes a universal Cartan, later to be identified with a Cartan subgroup of $\mathfrak{g}$ we let $M_\lambda$ denote the module with the universal property that for any $\mathfrak{g}$-module:

$$\mathrm{Hom}_{\mathfrak{g}}(M_\lambda, V) = \mathrm{Hom}_{\mathfrak{b}}(\mathbb{C}_\lambda, V).$$

The module $\mathfrak{M}_\lambda$ is called the *Verma module* of weight $\lambda$, and it is very easy to see that it exists, namely:

$$M_\lambda = U(\mathfrak{g}) \otimes_{U(\mathfrak{b})} \mathbb{C}_{\lambda-\rho}.$$

[1] In class, but not in the notes!

Notice that, by the PBW theorem, as a $\mathfrak{b}^-$-module:

$$M_\lambda = U(\mathfrak{n}^-) \otimes_{\mathbb{C}} \mathbb{C}_{\lambda-\rho}, \tag{17.1}$$

where $U(\mathfrak{n}^-)$ acts by left multiplication on the first factor, and the $\mathfrak{h}$-action is the tensor product of the adjoint representation and the representation on $\mathbb{C}_{\lambda-\rho}$. (The tensor product of Lie algebra representations is defined as: $X(v \otimes w) = (Xv) \otimes w + v \otimes (Xw)$.)

Therefore:

**Lemma 1.1.** *1. $M_\lambda$ is $\mathfrak{h}$-locally finite and semisimple. The ($\mathfrak{h}$-)weights of $M_\lambda$ are of the form $(\lambda-\rho) - \sum_i c_i\alpha_i$, where $\alpha_i$ range over simple positive roots (we will denote their set by $\Delta$) and $c_i \in \mathbb{Z}$. The weight spaces are finite-dimensional, and $M_\lambda^{\lambda-\rho}$ is one-dimensional.*

*2. $M_\lambda$ is $\mathfrak{n}$-locally finite.*

*Proof.* The first statement follows immediately from the presentation (**??**), and the second from the first and the fact that the action of $\mathfrak{n}$ raises weights. □

**Corollary 1.2.** *$M_\lambda$ has a unique irreducible quotient, which will be denoted by $L_\lambda$.*

*Proof.* The sum of all proper submodules does not meet $M_\lambda^{\lambda-\rho}$, and hence is proper. □

## 17.2 The category $\mathcal{O}$.

The full subcategory of $\mathfrak{g}$-modules which are:

- $\mathfrak{h}$-locally finite and semisimple;
- $\mathfrak{n}$-locally finite;
- finitely generated

is called category $\mathcal{O}$ (from a Russian paper of Gelfand-Gelfand-Bernstein). As we have seen, it contains Verma modules.

**Lemma 2.1.** *A submodule of a module in $\mathcal{O}$ is in $\mathcal{O}$. The category is* noetherian*, i.e. every increasing chain of subobjects of a given object stabilizes.*

*Proof.* For the first statement, only finite generation is not obvious, but it follows from the fact that $U(\mathfrak{g})$ is noetherian (a corollary of PBW).

The union of a chain of submodules is a submodule, hence finitely generated, hence the chain has to stabilize. □

We will eventually see that it is also Artinian, i.e. every object is of finite length.

**Lemma 2.2.** *Every object in $\mathcal{O}$ has a filtration whose quotients are surjective images of Verma modules.*

*Proof.* Let $V$ be in $\mathcal{O}$, and let $W \subset V$ be a finite-dimensional, generating subspace. Without loss of generality, $W$ is $\mathfrak{b}$-stable (for $U(\mathfrak{b})W$ is, in any case, finite-dimensional). By Lie's theorem, it has a filtration with one-dimensional quotients. Therefore, $V$ has a filtration with quotients generated by $\mathfrak{b}$-eigenvectors. Each such representation is the surjective image of a Verma module. □

The *Grothendieck group* of an abelian category $\mathcal{C}$ is the free group on its objects, modulo the relation: $[B] = [A] + [C]$ for every short exact sequence $0 \to A \to B \to C \to 0$. We will eventually see that the Grothendieck group of $\mathcal{O}$ is generated by Verma modules, in fact: it is free on the set of Verma modules.

## 17.3 The case of $\mathfrak{sl}_2$, and application.

Here we identify the elements of $\mathfrak{h}^*$ with integers, according to their value on $\check{\alpha}$. Under this, $\rho$ corresponds to 1, and therefore $M_\lambda$ denotes the Verma module with heighest weight $\lambda - 1$.

**Lemma 3.1.** *$M_\lambda$ is irreducible, unless $\lambda \in \mathbb{Z}_{>0}$, in which case there is an exact sequence:*

$$0 \to M_{-\lambda} \to M_\lambda \to L_\lambda \to 0.$$

*Proof.* Every submodule must have a highest weight vector, which must be of the form $F^n v_{\lambda-\rho}$. We compute that:

$$EF^n v_{\lambda-\rho} = n(\lambda - n)F^{n-1} v_{\lambda-\rho},$$

therefore for it to be zero (for some $n > 0$ we must have $\lambda \in \mathbb{Z}_{>0}$. □

We return to the case of a general semisimple $\mathfrak{g}$. Then:

**Lemma 3.2.** *If $\alpha$ is a simple root such that $\langle \lambda, \check{\alpha} \rangle \in \mathbb{Z}_{>0}$ then there is an embedding: $M_{s_\alpha \lambda} \hookrightarrow M_\lambda$. The quotient $V = M_\lambda / M_{s_\alpha \lambda}$ has the property that it is locally $(\mathfrak{sl}_2)_\alpha$-finite, where $(\mathfrak{sl}_2)_\alpha$ denotes the embedding of $\mathfrak{sl}_2$ into $\mathfrak{g}$ determined by the root $\alpha$.*

*Proof.* As in the previous lemma, we calculate that there is a highest weight vector with weight $s_\alpha \lambda$, hence there is a non-trivial map: $M_{s_\alpha \lambda} \to M_\lambda$. Since $M_{s_\alpha \lambda}, M_\lambda \simeq U(\mathfrak{n}^-)$ as $U(\mathfrak{n}^-)$-modules, and $U(\mathfrak{n}^-)$ does not have zero divisors, such a map has to be injective.

With notation $(H_\alpha, E_\alpha, F_\alpha)$ for the $\mathfrak{sl}_2$-triple corresponding to $\alpha$, we need to show that the quotient is $F_\alpha$-locally finite. (Finiteness under the other two is automatic for the category $\mathcal{O}$.) If $V'$ is the set of $F_\alpha$-finite vectors, then $V' \ni v_{\lambda-\rho}$; we claim that $V'$ is $\mathfrak{g}$-stable. Indeed, we have a homomorphism of $F_\alpha$-modules: $\mathfrak{g} \otimes V' \to V$, where $F_\alpha$ acts on $\mathfrak{g}$ via the adjoint representation. But $\mathfrak{g}$ is $F_\alpha$-finite and $V'$ is $F_\alpha$-locally finite, hence their tensor product is locally finite, therefore $\mathfrak{g}V' \subset V'$. Together with $v_{\lambda-\rho} \in V'$, this implies that $V' = V$. □

We haven't defined characters yet, but here's a corollary about them:

**Corollary 3.3.** *For every subquotient of $V$, the character is $s_\alpha$-stable.*

This says, in particular, that the set of weights of that subquotient is $s_\alpha$-stable.

The corollary follows from the theory of finite-dimensional $\mathfrak{sl}_2$-representations. It implies:

**Corollary 3.4.** *Assume that $\lambda$ is* integral *(i.e. $\langle\check{\alpha},\lambda\rangle \in \mathbb{Z}$ for all roots $\alpha$) and* strictly dominant *(i.e. $\langle\check{\alpha},\lambda\rangle > 0$ for all positive roots $\alpha$). Then the represenation:*

$$L'_\lambda = M_\lambda / \left(\sum M_{s_\alpha \lambda}\right)$$

*(sum over simple positive roots) is finite dimensional. In particular, $L_\lambda$ (the unique irreducible quotient of $M_\lambda$) is an (the) irreducible finite-dimensional representation with highest weight $\lambda - \rho$.*

*Proof.* By the previous corollary, the quotient will have a $W$-stable set of weights. On the other hand, all weights are $\leqslant \lambda$ and differ from $\lambda$ by an element of the root lattice, so there is a finite set of weights only. Finally, the weight spaces are finite dimensional, so the quotient is finite-dimensional. □

We will eventually see that $L'_\lambda = L_\lambda$. In particular, for each dominant integral weight $\lambda$ the *exists* a (unique up to isomorphism) finite-dimensional representation $V_\lambda := L_{\lambda+\rho}$ of $\mathfrak{g}$. (The uniqueness was a corollary of Corollary **??**.)

## 17.4 Localization with respect to $\mathfrak{z}(\mathfrak{g})$

Recall that $\mathfrak{z}(\mathfrak{g})$ denotes the center of the universal enveloping algebra. Lemma **??** implies:

**Lemma 4.1.** *$\mathfrak{z}(\mathfrak{g})$ acts by scalars on Verma modules. For every object $V$ in $\mathcal{O}$, the action of $\mathfrak{z}(\mathfrak{g})$ on $V$ is locally finite.*

*Proof.* For the first statement, notice that $\mathfrak{z}(\mathfrak{g})$ preserves weight spaces, so it must act by a scalar on the one-dimensional highest weight space. But that generates the whole module, so it acts by the same scalar on it.

We have seen in Lemma **??** that every object can be filtered by surjective images of Verma modules. The center acts by a scalar on a Verma module, hence on its quotients. Therefore it acts locally finitely on finite extensions of such objects. □

The following will be proven later:

**Theorem 4.2** (Harish-Chandra isomorphism)**.** *There is an isomorphism of algebras $\phi : \mathfrak{z}(\mathfrak{g}) = \mathbb{C}[\mathfrak{h}^*]^W$, with the property that every $\Delta \in \mathfrak{z}(\mathfrak{g})$ acts on the Verma module $V_\lambda$ by the scalar $\phi(\Delta)(\lambda)$.*

Notice that the maximal ideals of $\mathbb{C}[\mathfrak{h}^*]^W$ are the complex points of the set-theoretic quotient $\mathfrak{h}^*/W$. Indeed, for every finite group $\Gamma$ acting on an affine variety $X$ the quotient $X/\!\!/\Gamma$ is an affine variety (i.e. $k[X]^\Gamma$ is finitely generated), the map $X \to X /\!\!/ \Gamma$ is finite, and the quotient is also the so-called *geometric quotient* (usually denoted $X/\Gamma$, although this can be mistaken for the *stack-theoretic* quotient, which is the most "correct" one), in particular its points are in bijection with $G$-orbits on $X$.

**Corollary 4.3.** *1. The category $\mathcal{O}$ is a direct sum of categories $\mathcal{O}_\chi$, with $\chi$ varying over the complex points of $\mathfrak{h}^*/W$.*

*2. If $\lambda$ is such that $\lambda - w\lambda$ is never a sum of the form $\sum_\alpha n_\alpha \alpha$, with $\alpha$ varying over simple (positive) roots and $n_\alpha \in \mathbb{N}$, then $M_\lambda$ is irreducible.*

*3. Every object in $\mathcal{O}$ is of finite length.*

*4. The classes of the Verma modules $M_\lambda$ (or, equivalently, their irreducible quotients $L_\lambda$) are a basis for the Grothendieck group $\mathbb{Z}[\mathcal{O}]$.*

*Proof.* 1. Since the action of the center is locally finite, we can decompose every object into a direct sum of generalized $\mathfrak{z}(\mathfrak{g})$-eigenspaces. Obviously, there are no $\mathfrak{g}$-morphisms between them.

2. If $M_\lambda$ is not irreducible, there is a nontrivial map from $M_\mu$ to $M_\lambda$ for some $\mu$. But this can happen only if $\lambda - \mu$ is in the positive root monoid, and by the decomposition of categories it can only happen if $\mu = w\lambda$ for some $w \in W$.

3. By Lemma **??**, it suffices to show that Verma modules are of finite length. Define $K_\lambda$ by the short exact sequence:

$$0 \to K_\lambda \to M_\lambda \to L_\lambda \to 0,$$

if nonzero then $K_\lambda$ admits a filtration as in Lemma **??**, whose factors are surjective images of modules $M_\mu$ with $\mu \lneq \lambda$. But by the decomposition of categories, $\mu$ has to be a $W$-conjugate of $\lambda$, and hence in a finite number of steps we will arrive at weights $\mu$ as in the previous statement, hence $M_\mu$ irreducible.

4. Same argument, by induction on $\lambda$ (the starting point of the induction being the $M_\lambda$ of 2. The fact that the $L_\lambda$ also form a basis follows from the fact that the category is Artinian, and they are the only irreducible objects (non-isomorphic to each other).

□

**Corollary 4.4.** *When $\lambda$ is integral and strictly dominant, the representations $L'_\lambda$ of the previous subsection are irreducible (hence equal to $L_\lambda$).*

## 17.5 Characters

Consider the ring $R$ of formal sums $\sum_{\lambda\in\mathfrak{h}^*} c(\lambda)e^\lambda$, where $c_\lambda$ is supported in a finite number of translates of the negative root monoid. Elements of the ring are multiplied as the notation suggests.

The *character* $\mathrm{ch}_V$ of an object $V$ in the category $\mathcal{O}$ (or a sub-$\mathfrak{h}$-module) is the following element of $R$: $\sum_{\lambda\in\mathfrak{h}^*} \dim V_\lambda e^\lambda$. It defines a group homomorphism from $\mathbb{Z}[\mathcal{O}]$ (the Grothendieck group of $\mathcal{O}$) to $R$. Moreover, it can easily be shown that the character of the tensor product of two representations is the product of the two characters. (The tensor product is not necessarily in $\mathfrak{O}$, but it has weight spaces such that this statement makes sense.)

Let $L = \prod_{\alpha>0}\left(e^{\frac{\alpha}{2}} - e^{-\frac{\alpha}{2}}\right) \in R$ (the product over all positive roots). Notice that $L$ can also be written as: $e^\rho \prod_{\alpha>0}\left(1-e^{-\alpha}\right)$, hence it is supported on *integral* weights. (The weight $\rho$ is integral because $\alpha = \rho - s_\alpha\rho = \langle\rho,\check{\alpha}\rangle\,\alpha$ for every simple root $\alpha$.)

**Proposition 5.1.** *The character of the Verma module $M_\lambda$ satisfies:*

$$L\cdot\chi_{M_\lambda} = e^\lambda.$$

*Proof.* As an $\mathfrak{h}$-module, $V_\lambda = U(\mathfrak{n}_-)\otimes\mathbb{C}_{\lambda-\rho}$, so $\mathrm{ch}(V) = \mathrm{ch}(U(\mathfrak{n}_-))\cdot\mathrm{ch}(\mathbb{C}_{\lambda-\rho}) = \chi(U(\mathfrak{n}^-))\cdot e^{\lambda-\rho}$. Therefore, it suffices to prove that the character of $U(\mathfrak{n}_-)$ is the power series $e^\rho/L$.

By the Poincare-Birkhoff-Witt theorem, as $\mathfrak{h}$ representations we have: $U(\mathfrak{n}_-) = \otimes_{\alpha>0} S^\bullet\mathfrak{g}_{-\alpha}$. The character of $S^\bullet\mathfrak{g}_{-\alpha}$ is $1+e^\alpha+e^{2\alpha}+\cdots = \frac{1}{1-e^{-\alpha}}$, and this proves the proposition. □

Finally, we are ready to prove the *Weyl character formula*:

**Theorem 5.2.** *The character of the irreducible representation with heighest weight $\lambda$ is given by the* Schur polynomial*:*

$$\chi(V_\lambda) = \frac{\sum_{w\in W}\mathrm{sgn}(w)e^{w(\lambda+\rho)}}{L}$$

*Proof.* Since in the Grothendieck group we have:

$$[V_\lambda] = [M_{\lambda+\rho}] + \sum_{w\in W, w\neq 1} c_w[M_{w(\lambda+\rho)}],$$

we get:

$$L\cdot\mathrm{ch}(V_\lambda) = e^{\lambda+\rho} + \sum_{w\in W, w\neq 1} c_w e^{w(\lambda+\rho)}.$$

On the other hand, we know that the character should be $W$-invariant (Corollary **??**), therefore the expression on the right should be $(W,\mathrm{sgn})$-equivariant. Therefore, $c_w = \mathrm{sgn}(w)$. □

# Chapter 18

# The Chevalley and Harish-Chandra isomorphisms

Later

# Chapter 19

# Commutative $C^*$- and Von Neumann algebras

We begin with a look at the theory of $C^*$- and Von Neumann (or $W^*$)-algebras. To every locally compact group $G$ one can associate the $C^*$-algebra $C^*(G)$, which is the $C^*$-*envelope* of the convolution algebra $L^1(G)$. General theorems about the structure of $C^*$- and $W^*$-algebras enable us, then, to decompose any unitary representation of $G$ into a direct integral of irreducibles (*Plancherel decomposition*) *provided* the $C^*$-algebra of the group is nice enough.

## 19.1 Basic definitions

A *Banach algebra* is a Banach space $(V, \| \bullet \|)$ with an algebra structure satisfying: $\|ab\| \leqslant \|a\| \cdot \|b\|$. Not to be confused with a $B^*$-algebra, which is a term synonymous to a $C^*$-algebra and now obsolete.

A *Banach *-algebra* is a Banach algebra equipped with an antiinvolution: $(ab)^* = b^*a^*$ satisfying: $\|a^*\| = \|a\|$.

A $C^*$ *algebra* is a Banach *-algebra which, in addition, satisfies $\|a^*a\| = \|a\|^2$. This turns out to be a very rigid condition: as we will see, the algebraic structure together with the requirement of completeness completely determine the norm.

Up to now our algebras were equipped with the norm topology. If, however, the space $V$ of a $C^*$-algebra turns out to be the Banach dual of a Banach space $V_*$ (called the *predual* of $V$), then we call it a $W^*$-*algebra* (or *Von Neumann algebra*) and we endow it with the weak-* topology. The weak-* topology has completely different features than the norm topology (for instance, the unit ball is compact, by the *Banach-Alaoglu theorem*), and this makes the study of Von Neumann algebras quite differnt. As we will see in this lecture, for commutative algebras this is the difference between topology and measure theory: (locally compact) Hausdorff spaces and (locally finite) measure spaces. Notice that the predual of a general Banach space is not uniquely defined: nonisomorphic spaces can have isomorphic duals (for example, the spaces $c$ and $c_0$ of convergent, resp.

nullsequences have duals isomorphic to the space $l^1$ of summable sequences). However, it turns out that for $W^*$ algebras the predual is unique.

Those algebras are called *unital* if they have an identity element. $W^*$-algebras turn out be unital always, $C^*$-algebras are not necessarily unital but they have an *approximate identity*: a net of elements $u_i$ with $\|u_i\| \leqslant 1$ and $\lim u_i a = \lim a u_i = a$ for all $a$. Even if a $C^*$ algebra $A$ does not have an identity, we can add one by defining on $A \oplus \mathbb{C}$ the norm:

$$\|\lambda 1 + a\| = \sup_{x \neq 0} \frac{\|\lambda x + ax\|}{\|x\|}.$$

However, since (two-sided) ideals in $C^*$-algebras are non-unital $C^*$-algebras, one needs to discuss non-unital $C^*$-algebras as well.

An example of a $C^*$-algebra is the algebra $B(H)$ of bounded operators on a Hilbert space, with $*$ : the adjoint operator. It turns out that this is a Von Neumann algebra:

**Theorem 1.1.** *Let $T(H)$ be the Banach space of* trace class operators *in $B(H)$; the norm on $T(H)$ is given by $\|a\|_{TC} = \|(a^*a)^{\frac{1}{4}}\|_{HS}$, where $HS$ is the Hilbert-Schmidt norm and $(a^*a)^{\frac{1}{4}}$ is defined by continuous functional analysis (will discuss later). Then we have:*

$$B(H) = T(H)^*$$

*with pairing $(a, b) = \mathrm{tr}(a^*b)$.*

It turns out that $C^*$- and $W^*$- algebras have faithful representations under which we can equivalently define them as (those algebras which are isomorphic to) *norm-closed*, respectively *weak-* closed* subspaces of $B(H)$, for some Hilbert space $H$. Moreover, the weak-* topology on $B(H)$ coincides with what is called the *weak operator topology*: a net of operators $T_i \to T$ if an only if for all $x, y \in H$ we have: $\langle T_i x, y\rangle \to T$.

## 19.2 Invertible elements and characters of a commutative $C^*$-algebra

From now on, for the rest of the lecture, all $C^*$-algebras are commutative.

An example of a commutative $C^*$-algebra is the space $C_0(X)$ of continuous functions vanishing at infinity on a locally compact space $X$, endowed with the supremum norm and the involution $f \mapsto \bar{f}$. It is unital if and only if $X$ is compact.

*From now for the rest of this section on all Banach algebras are unital with*[1] $\|I\| = 1$.

[1] This condition is automatic for $C^*$-algebras. For a general Banach algebra $A$, the given norm is equivalent to the norm $a \mapsto \|L_a\|$, where $L_a \in B(A)$ is the operator of left multiplication by $a$, and for that we have $\|I\| = 1$.

The converse is true: every $C^*$-algebra $A$ turns out to be isomorphic to $C_0(X)$ for some locally compact space $X$. To recover $X$ from $A$, we think as in algebraic geometry: the points of $X$ should be maximal ideals of $A$, or equivalently homomorphisms into a field.

The following hold automatically:

**Proposition 2.1.** 1. *Every (algebraic) maximal ideal in a Banach algebra is closed.*

2. *The only (complex) Banach field is $\mathbb{C}$. Hence, maximal ideals of $A$ are in bijection with* characters*, that is homomorphisms: $A \to \mathbb{C}$ (automatically continuous, since the ideal is closed).*

3. *For every character $\omega$ we have $\|\omega\| = 1$.*

4. *Every maximal ideal in a $C^*$-algebra is self-adjoint, i.e. closed under *. (This turns out to be true for any closed ideal, but it's harder and we won't need it.) Hence, every character is a *-homomorphism into $\mathbb{C}$.*

For the proof, and for later use, we introduce the *spectrum* $\sigma(a)$ of an element $a \in A$: it is the set of $\lambda \in \mathbb{C}$ such that $a - \lambda I$ is not invertible. We will prove the above proposition together with the following:

**Proposition 2.2.** 5. *The spectrum of every element in a Banach algebra is a nonempty compact subset of $\mathbb{C}$.*

*Proof.* 1. We start with a lemma. For what follows, keep in mind

**Lemma 2.3.** *If $\|a\| < 1$ then the series $\sum_{n=0}^{\infty} a^n$ converges to $(1-a)^{-1}$.*

The proof is easy. It has two corollaries:

a. If $\lambda > \|a\|$ then $a - \lambda I$ is invertible. Indeed, $\|\lambda^{-1}a\| < 1$.

b. The set of invertible elements is open. Indeed, if $a$ is invertible and $\|b\| < \|a^{-1}\|^{-1}$ then $\|a^{-1}b\| < 1$ and hence $a + b = a(1 + a^{-1}b)$ is invertible.

Hence we get:

- Every maximal ideal is closed. For it doesn't contain any invertible elements, and since those are open neither does its closure, i.e. its closure is also a proper ideal.
- For each $a \in A$, $\sigma(a)$ is closed and bounded (compact).

5. To prove that the spectrum of an element $a \in A$ is nonempty, we define a function $f : \mathbb{C} \setminus \sigma(a) \to A$ by $g(z) = (zI - a)^{-1}$. We claim that this function is holomorphic,[2] and $\lim_{z\to\infty} g(z) = 0$. By Liouville's theorem,

[2]The notion of a *holomorphic* function into a Banach space, defined by the convergence of the usual derivative, is known to be equivalent to the notion of a *weakly holomorphic* function, which means that $\langle g(z), a^* \rangle$ is holomorphic for every $a^* \in A^*$.

if $\sigma(a) = \varnothing$ this will imply that $\langle g(z), a^* \rangle = 0$ for every $a^* \in A^*$, which is absurd.

Holomorphicity is proven by showing that in an open ball around a point $z_0$ which doesn't meet the spectrum we have:

$$g(z) = \sum_{n=0}^{\infty} (z_0 - z)^n (z_0 I - a)^{-n-1},$$

a convergent power series. The fact that $g(z) \to 0$ as $z \to \infty$ follows from writing:

$$\|g(z)\| = |z|^{-1} \|(I - z^{-1}a)^{-1}\|$$

and observing that $\lim_{z\to\infty}(I - z^{-1}a) = I$.

2. If $A$ is a Banach field, $a \in A$ and $z \in \sigma(z)$ then $a - zI$ is not invertible, hence zero. This means that $a = zI$, and $A = \mathbb{C}$.

3. We have seen that if $z > \|a\|$ then $a - zI$ is invertible and hence $\omega(a - zI) \neq 0$. Therefore, $|\omega(a)| \leqslant \|a\|$, but $\omega(I) = 1 = \|I\|$, so $\|\omega\| = 1$.

4. It suffices to show that for a character $\omega$, and a self-adjoint element $a$ (i.e. $a = a^*$, we have $\omega(a) \in \mathbb{R}$. Assume $\omega(a) = \alpha + i\beta$, and by replacing $a$ by the self-adjoint element $a - \alpha I$ we may assume that $\omega(a) = i\beta$. Then we compute that for $t \in \mathbb{R}$:

$$|\omega(a + itI)|^2 = \beta^2 + 2t\beta + t^2.$$

On the other hand, $\|a + itI\|^2 = \|a^2 + t^2 I\| \leqslant \|a\|^2 + t^2$, and if $\beta \neq 0$ and $t \gg 0$ we get a contradiction to the fact that $\|\omega\| = 1$.

□

## 19.3 The Gelfand transform

Given a commutative $C^*$-algebra $A$ (or more generally any Banach algebra), we will denote its set of characters by $\Delta(A)$ and we will call it the *spectrum*[3] or *structure space* of $A$. If $A$ has a unit element then $\Delta(A)$ is weak-* closed in $A^*$ (if it doesn't we need to include the zero homomorphism). We endow $\Delta(A)$ with the weakest topology making all elements of $a$ continuous, that is: the restriction of the weak-* topology when $\Delta(A)$ is considered as a subspace of $A^*$. The Banach-Alaoglu theorem implies:

*If $A$ has an identity element then $\Delta(A)$ is a compact (Hausdorff) space.*

In the general case, it is a locally compact space. The map $A \to C_0(\Delta(A))$: $a \mapsto \hat{a}(\omega) := \omega(a)$ is the *Gelfand transform*. (The reason that functions vanish at infinity is that "infinity" corresponds to the zero functional.) Proposition **??** implies:

[3]What we called "spectrum" $\sigma(a)$ of an element $a \in A$ before becomes the range of its Gelfand transform $\hat{a}$, as a function on $\Delta(A)$.

**Corollary 3.1.** *The Gelfand transform is a *-homomorphism.*

The basic goal of this lecture is to prove the following theorem, which shows that *all* commutative $C^*$-algebras are of the form $C_0(X)$, for some locally compact space $X$:

**Theorem 3.2.** *For a commutative $C^*$-algebra $A$ the Gelfand transform is an isometric isomorphism onto $C_0(\Delta(X))$.*

It is enough to prove the theorem for unital $C^*$-algebras; the general case follows by adding an identity element and extending characters (this can be done in a unique way). So let $A$ be a unital $C^*$-algebra.

Theorem **??** will be proven by showing:

**Proposition 3.3.** *The map $A \ni a \mapsto \hat{a} \in C(\Delta(A))$ is an isometry.*

This shows that the Gelfand transform is injective, and by an easy application of the Stone-Weierstrass theorem it can be seen to be surjective.

To prove it we need to have a more intrinsic, "algebraic" definition of the norm.

**Proposition 3.4.** *1. For a self-adjoint element $a$ (i.e. $a^* = a$) we have $\|a\| = \sup|\sigma(a)|$.*

**Lemma 3.5.** *We have $\sigma(a) = \sigma(\hat{a})$.*

*Proof.* An element $a$ is invertible iff it belongs to no maximal ideal iff $\omega(a) \neq 0$ for all $\omega \in \Delta(X)$. Thus, $a - zI$ is invertible iff $\hat{a} - z$ does not vanish (is invertible). □

This shows how Proposition **??** follows from Proposition **??**: $\|a\| = \sup|\sigma(a)| = \sup|\sigma(\hat{a})| = \|\hat{a}\|$ for all self-adjoint elements, which is enough in order to prove it for all elements (just apply the equality to the element $aa^*$).

We are left with proving Proposition **??**. The number $r(a) = \sup|\sigma(a)|$ is called the *spectral radius* of $a$.

**Proposition 3.6** (Spectral radius formula)**.** *For any element $a$ in a unital Banach algebra we have:*

$$r(a) = \lim \|a^n\|^{\frac{1}{n}}.$$

This proves Proposition **??**, since for self-adjoint elements $\|a^{2^n}\| = \|a\|^{2^n}$.

*Proof.* I leave it as an exercise to prove that for a polynomial $p$, $p(\sigma(a)) = \sigma(p(a))$. Therefore, $r(a)^n = r(a^n) \leqslant \|a^n\| \Rightarrow r(a) \leqslant \|a^n\|^{\frac{1}{n}}$, for all $n$.

On the other hand, the function $g(z) = (zI - a)^{-1}$ that we encountered in the proof of Proposition **??** is analytic for $|z| > r(a)$ and has a convergent expression:

$$g(z) = \frac{1}{z}\sum_{n=0}^{\infty}\left(\frac{a}{z}\right)^n$$

when $|z| > \|a\|$. Therefore, the same expression should converge, locally uniformly in $z$, when $|z| > r(a)$. This means, in particular, that $\left(\frac{a}{z}\right)^n$ is bounded, hence $\limsup \|a^n\|^{\frac{1}{n}} \leqslant |z|$. Since this holds for every $z$ with $|z| > r(a)$, the claim is proven. □

This completes the proof of Theorem **??**.

## 19.4 Commutative $W^*$-algebras

Now let us see the analogous theorem for $W^*$-algebras. Here the locally compact Hausdorff space $\Delta(A)$) will be replaced by a measure space $(X, \Omega, \nu)$ which is *locally finite*, i.e. a direct sum of finite measure spaces. For such a space, the commutative $C^*$-algebra $L^\infty(X, \nu)$ is actually a $W^*$-algebra, being the dual of $L^1(X, \nu)$. Vice versa:

**Theorem 4.1.** *Let $A$ be a commutative $W^*$-algebra. Then it is isomorphic to $L^\infty(X, \nu)$ for some locally finite measure space $(X, \Omega, \nu)$.*

The proof of this theorem should be read after the next chapter.

*Proof.* By the Gelfand-Naimark theorem, $A = C(X)$, where $X$ is the spectrum of $A$. If $\phi$ is a normal state then by the Riesz representation theorem it defines a unique Radon measure $\mu_\phi$ on $X$, easily seen to be positive, such that:

$$\phi(a) = \int_X \hat{a}\mu_\phi.$$

This gives a morphism $\iota : A \to L^\infty(X, \mu_\phi)$ which is easily seen to be a $W^*$-morphism (i.e. a weak-star continuous *-homomorphism): Indeed, it is immediate that it is a *-homomorphism. For weak-star continuity: first, a weak-* convergent net $x_\alpha \to x$ in $A$ is bounded (by the uniform boundedness principle). Since $\iota$ is norm-continuous, $(\iota x_\alpha)_\alpha$ is bounded. Therefore, it suffices to show that $\langle f, \iota x_\alpha\rangle \to \langle f, \iota x\rangle$ for $f$ in a dense subset of $L^1(X, \mu_\phi)$ – we take $f$ to be continuous, hence $f = \hat{b}$ for some $b \in A$. Then $\langle f, \iota x_\alpha\rangle = \int \hat{b}\hat{x}_\alpha\mu_\phi = \phi(bx_\alpha) \to \phi(bx)$ since $\phi$ is weak-star continuous.

We now know that the image of $A$ is weak-star closed, and since it contains the image of continuous functions it is equal to $L^\infty(X, \mu_\phi)$.

The kernel of $\iota = \iota_\phi$ is a closed ideal, and hence generated by a projection $z \in A$. It is easily seen that $z = I - \mathrm{supp}(\phi)$.

□

# Chapter 20

# General $C^*$-algebras and their states

## 20.1 Corollaries of the Gelfand-Naimark theorem

The Gelfand-Naimark theorem can often be used to deduce results for arbitrary $C^*$-algebras, as follows: if $a \in A$ is a *normal* element, i.e. commutes with its adjoint: $aa^* = a^*a$, then it generates a commutative $C^*$-algebra inside of $A$.

**Theorem 1.1.** *Let $\iota : A \hookrightarrow B$ be an injective homomorphism from a $C^*$-algebra to a normed algebra. Then $\|x\|_A \leqslant \|\iota x\|_B$ for any normal element $x$; if $\iota$ is an injective $^*$-homomorphism of $C^*$ algebras, it is an isometry.*

*Proof.* By considering the $C^*$-algebra generated by $x$, we may first assume that $A$ is commutative, hence without loss of generality $B$ is also commutative. If $X$ is the spectrum of $A$ and $X' \subset X$ is the set of characters which are $B$-continuous, then we claim that $X'$ is dense in $X$. Indeed, if not we can find an open subset $U$ of $X$ and nonzero elements $a, b \in A$ such that $\hat{a}$ is supported in $U$ and $\hat{b}$ is 1 on $X'$ and 0 on $U$. Hence, $ab = 0$.

On the other hand, we claim that $\iota(b)$ is invertible in $B$; indeed, otherwise there is a character $\omega$ of $B$ with $\omega(\iota(b)) = 0$, but by assumption $\omega(b) = 1$ for every character of $B$. Therefore, $\iota(ab) = 0$ implies that $\iota(a) = 0$, contradicting the injectivity of $\iota$.

Now recall that $\|x\|_A = r_A(x)$ and $\|x\|_B \geqslant r_B(x)$, where by $r$ we denote the corresponding spectral radii. Since $X'$ is dense in $X$, it follows that $r_A(x) = r_B(\iota(x))$, from which we deduce the first statement. If $B$ is also a $C^*$-algebra, the inequality becomes an equality, and therefore the map is an isometry.

For an injective $^*$-homomorphism $A \hookrightarrow B$ of $C^*$-algebras (not necessarily commutative), the norms are determined by those of normal elements (by the formula $\|x\|^2 = \|x^*x\|$), and therefore the general statement follows from the one about commutative algebras. □

**Corollary 1.2.** *The image of a *-homomorphism of C*-algebras is always closed.*

**Theorem 1.3.** *If $\iota : A \to B$ is a morphism of W*-algebras (i.e. a *-homomorphism which is continuous under the weak-* topologies) then the image is closed.*

*Proof.* First, observe that a weak-star continuous operator between Banach spaces is bounded; in fact, it is the adjoint of a bounded operator between the preduals. (A weakly continuous operator between Banach spaces $T : V \to W$ is bounded by applying the *uniform boundedness principle* to $W^*$: for a norm-convergent sequence $x_n \to 0$ in $V$ the functionals $\phi \mapsto \phi(Tx_n)$ on $W^*$ are all bounded, and hence uniformly bounded.)

Hence, by using the analogous statement for $C^*$-algebras, $\iota$ is an isometry from $A/K$ to $B$, where $K$ is the kernel, and its image is norm-closed. Since the image of the unit ball in $A/K$ is the unit ball in $\iota(A)$, it suffices to show that this is weak-star closed. But it is the continuous image of a compact set (in the weak-star topology), hence compact, hence closed. □

## 20.2 Positive elements

Let $A$ be a $C^*$-algebra. An element $a \in A$ is called *positive* if it is self-adjoint, and its spectrum is contained in $[0, +\infty)$. Hence, under the Gelfand transform for the $C^*$-algebra that it generates, it corresponds to a non-negative real-valued function. We write $a > 0$ if $a$ is positive, and $a < 0$ if $-a > 0$; the set of positive elements is denoted by $A^+$. By considering the Gelfand transform it immediately follows:

**Lemma 2.1.** *A self-adjoint element $a$ is positive if $\left\| 1 - \frac{a}{\|a\|} \right\| \leqslant 1$.*

**Theorem 2.2.** *1. Positive elements form a closed cone.*

*2. An element $a$ is positive if and only if there exists a $x \in A$ such that $x^*x = a$. There exists a unique positive such $x$.*

*Proof.* The first statement follows easily by using the previous lemma.

For the second, if $a$ is positive the existence of $x$ is immediate from the Gelfand-Naimark theorem. Vice versa, if $a = x^*x$ but $a$ is not positive then (again by the same theorem) there exists a positive element $b$ such that $bx^*xb < 0$. By writing $xb = h_1 + ih_2$, with $h_i$ self-adjoint, we compute that $(xa)^*(xa) + (xa)(xa)^* = 2(h_1^2 + h_2^2) > 0$. (We are using here the first fact, that the sum of two positive elements is positive.)

On the other hand, the spectrum of $(xa)^*(xa)$ union $\{0\}$ coincides with the spectrum of $(xa)(xa)^*$ union $\{0\}$ as the following lemma shows. Hence, they are both negative, and their sum is negative, a contradiction.

For uniqueness of such a positive $x$, it suffices to show that such a $x$ should commute with $a$ (then it follows again from Gelfand-Naimark); but this is obvious since $a = x^*x = x^2$. □

**Lemma 2.3.** *If $a, b$ are two elements in a $C^*$-algebra, $s(ab) \cup \{0\} = s(ba) \cup \{0\}$.*

*Proof.* Assume that $ab - zI$ is invertible, $= u^{-1}$. Then we compute: $(ba - zI)(bua - I) = zI = (bua - I)(ba - zI)$. Hence, if $z \neq 0$ then $ba - zI$ is invertible. □

*Remark.* Unlike the finite-dimensional case, we cannot avoid adjoining 0. Indeed, consider a Hilbert space with a proper isomorphic subspace, $\mathcal{H} = \mathcal{H} \oplus C$. If $a \in B(\mathcal{H})$ is projection to the subspace $\mathcal{H}$ and $b \in B(\mathcal{H})$ is the embedding of the subspace into the ambient space, then $ab$ is the identity, but $ba$ has nontrivial kernel.

Every self-adjoint element can be written as $a = a_+ - a_-$ with $a_\pm > 0$, commuting with each other, and $\max\{\|a_\pm\|\} = \|a\|$ (from the Gelfand-Naimark theorem); the element $|a| = a_+ + a_-$ is called the *absolute value* of $a$.

## 20.3 Positive functionals

The space $P$ of *positive functionals* on $A$ is the space of all linear functionals: $A \to \mathbb{C}$ which are $\geqslant 0$ on positive elements. A priori, we do not need to assume that they are bounded, but they will turn out to be.

As an example, on $B(\mathcal{H})$ (bounded operators on a Hilbert space) consider the functional $a \mapsto \langle a\xi, xi\rangle$, where $\xi \in \mathcal{H}$. It will actually turn out that all positive functionals of a $C^*$-algebra $A$ are of this form, for some representation $A \to B(\mathcal{H})$.

**Theorem 3.1.** *1. A positive functional is always bounded, hence $P \subset A^*$.*

*2. If $\phi \in P$ then:*

*(a) $\phi(a^*) = \overline{\phi(a)}$, i.e. it is a *-homomorphism into $\mathbb{C}$;*

*(b) the form $(a, b) \mapsto \phi(b^*a)$ is hermitian positive semi-definite;*

*(c) $|\phi(a)|^2 \leqslant \|\phi\|\phi(a^*a)$.*

*Proof.* First of all, if $\phi$ is unbounded then it is unbounded on self-adjoint elements (because $a = h_1 + ih_2$ with $h_i$ self-adjoint of norm $\leqslant \|a\|$); and hence it is unbounded on positive elements (because of the decomposition $a = a_+ - a_-$ with $\|a_\pm\| \leqslant \|a\|$).

Assume $a = \sum 2^{-n} a_n$ with $a_n > 0$, $\|a_n\| = 1$, $\phi(2^{-n}a_n) > 1$. Then $\phi(a) > N$ for every $N$, a contradiction. Hence $\phi$ is bounded.

To prove that $\phi$ is a *-homomorphism, we first show that it takes real values on self-adjoint elements. This follows from the decomposition $a = a_+ - a_-$. Now, for arbitrary $a = b + ic$ (with $b, c$ self-adjoint) we have $a^* = b - ic$ and hence $\phi(a^*) = \phi(b) - i\phi(c) = \overline{\phi(b + ic)}$.

Because of this, the form $(a, b) \mapsto \phi(b^*a)$ is clearly hermitian (i.e. sesquilinear, skew-symmetric), and it is positive since $\phi$ is positive.

Finally, from Cauchy-Schwarz inequality for this hermitian form (at least when $A$ has an identity; the general case can be handled with an approximate identity) we get: $\|\phi(a)\|^2 \leqslant \phi(1)\phi(a^*a) \leqslant \|\phi\|\phi(a^*a)$. □

*Remark.* More generally, an operator $\phi : A \to B$ between $C^*$-algebras is called *positive* if $\phi(A^+) \subset B^+$; *-homomorphisms are always positive.

## 20.4 States

A *state* of $A$ is a positive linear functional $\phi$ with $\|\phi\| = 1$; their space will be denoted by $\mathscr{S}$. Since we secretly know (and will see soon) that they correspond to symmetric matrix coefficients of representations, it is important to know that there are enough of them to distinguish points:

**Theorem 4.1.** *If $a \in A$ is normal then there is a $\phi \in \mathscr{S}$ with $|\phi(a)| = \|a\|$.*

*Proof.* There is certainly such a state (in fact, a character) on the $C^*$-algebra generated by $a$ (namely, evaluation at the maximum of $|\hat{a}|$). By the Hahn-Banach theorem, we may extend it to a linear functional on $A$ without changing its norm. The fact that the extension is positive follows from the following lemma (when $A$ has an identity; the general case is treated similarly with approximate identities):

**Lemma 4.2.** *A functional $\phi \in A^*$ is positive iff $\|\phi\| = \phi(1)$.*

Indeed, let $a$ be self-adjoint with $\|a\| \leqslant 1$; we first show that $\phi(a)$ is real. Without loss of generality, $\phi(a) = \alpha + i\beta$ with $\beta \leqslant 0$. Hence, $|\phi(a - inI)|^2 \leqslant 1 + n^2$ for every $n$; on the other hand, this is equal to $|\alpha + i\beta - in|^2 \geqslant (|\beta| + n)^2$, from which it follows that $\beta = 0$.

If $a$ is also positive then $\|I - a\| \leqslant 1 \Rightarrow 1 - \phi(a) \leqslant 1 \Rightarrow \phi(a) \geqslant 0$. This completes the proof of the lemma, and the theorem. □

## 20.5 Positivity and normal states for $W^*$-algebras

**Lemma 5.1.** *If $A$ is a $W^*$-algebra, the set of $A^s$ self-adjoint elements and the set $A^+$ of positive elements are both weak-star closed.*

*Proof.* It sufficies (Krein-Smulian theorem) to show that their intersection with the unit ball $A_1$ is closed. If $(x_\alpha)\alpha$ is a net of self-adjoint elements of norm $\leqslant 1$ converging in the weak-star topology to $a + ib$, assume $b \neq 0$ and that there is a positive $\lambda \in \sigma(b)$. (If not, consider the negative of this sequence.) Then $\|x_\alpha + inI\| \leqslant \sqrt{1 + n^2}$ and for large $n$ this is $< \lambda + n \leqslant \|b + nI\| \leqslant \|a + ibI + inI\|$. Since $a + ibI + inI$ is the limit of $(x_a + inI)_\alpha$ which belongs to the compact ball $\sqrt{1 + n^2}A_1$, this also belongs to that ball, a contradiction.

Recall that $A^+ \cap A_1$ contains precisely those self-adjoint elements $a \in A_1$ with $1 - \alpha \in A_1$, hence it is weak-star closed as the intersection of weak-star closed sets. □

**Corollary 5.2.** *The involution $*$ is weak-star continuous.*

A state on $A$ is called *normal* if it is weak-star continuous. Their set will be denoted by $\mathscr{S}_n$.

**Proposition 5.3.** *If $\phi(a) = 0$ for all $\phi \in \mathscr{S}_n$, then $a = 0$.*

*Proof.* By the Hahn-Banach separation theorem for *real* topological vector spaces: Say $a \in A^s \setminus A^+$ (otherwise work with $-a$). Since $A^+$ is a weak-star closed convex subset of the locally convex real topological vector space $A^s$, there is a real-valued linear functional $l$ on $A^s$ with $l(a) < \inf_{x \in A^+} l(x)$. Since $A^+$ is a cone, this means that $l(a) < 0$ and $l(A^+) \geqslant 0$. Now define $\phi(a+ib) = l(a)+il(b)$ on $A$ ($a, b \in A^+$). It is a continuous state with $\phi(a) \neq 0$. □

## 20.6 Universal representations

Let $A$ be a $C^*$-algebra. Given a state $\phi$, we can define the Hilbert space $\mathcal{H}_\phi$ as the completion of $A$ with respect to the seminorm $\|a\|_\phi = (\phi(a^*a))^{\frac{1}{2}}$. Left multiplication by $A$ is well-defined and gives a morphism $\pi_\phi : A \to B(\mathcal{H}_\phi)$, as is implied by the following lemma:

**Lemma 6.1.** *The radical of the hermitian form $(a, b) \mapsto \phi(b^*a)$ is a closed left ideal of $A$, and more precisely: $\|ba\|_\phi \leqslant \|a\|_\phi \cdot \|b\|$.*

*Proof.* Since the form is positive semi-definite, $\phi(a * a) = 0 \iff \forall b, \phi(b^*a) = 0 \Rightarrow \phi(a^*b^*ba) = 0$, so $ba$ is also in the radical. It is clearly closed by continuity of $\phi$.

To prove the norm estimate (which actually makes the first part of the proof redundant), for given $a$ with $\|a\|_\phi \neq 0$ we define a new functional: $\psi(b) = \phi(a^*ba)$. It is clearly positive, and hence $\psi(b) = \psi(I) = \|a\|_\phi^2$. Hence, $\psi(b^*b) \leqslant \|a\|_\phi^2 \cdot \|b\|^2$. (Use approximate identity if $I \notin A$.) □

Notice that if $\xi \in \mathcal{H}_\phi$ denotes the image of $I \in A$, we have: $\phi(a) = \langle \pi_\phi(a)\xi, \xi \rangle$.

We now form the huge Hilbert space: $\mathcal{H}_{\text{univ}} = \oplus_{\phi \in \mathscr{S}} \mathcal{H}_\phi$. It is called *the universal representation* of the $C^*$-algebra $A$. Theorem **??** shows that the universal representation $\pi_{\text{univ}} : A \to B(\mathcal{H}_{\text{univ}})$ is injective, and hence an isometry to its image. Therefore, *every $C^*$-algebra is isomorphic to a norm-closed $^*$-subalgebra of $B(\mathcal{H})$* (and vice versa, of course).

We now turn to $W^*$-algebras; assume that $A$ is such.

**Lemma 6.2.** *If $\phi$ is a normal state, then the map $\pi_\phi : A \to B(\mathcal{H}_\phi)$ is weak-star continuous.*

*Proof.* We use the following fact: the weak-star topology on $B(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space, is equivalent on bounded spheres to the *weak operator topology*, defined by the seminorms $T \mapsto |\langle T\xi, \eta \rangle|$ ($\xi, \eta \in \mathcal{H}$); in other words, the topology where $T_\alpha \to T$ iff $T_\alpha \xi \to T\xi$ weakly for all $\xi$. Since $\pi_\phi$ is norm-bounded, it suffices to prove continuity on bounded sets. (Indeed, it is easy to show that any weakly convergent net is norm-bounded.) In other words, we need to prove that for every $\xi, \eta \in \mathcal{H}_\phi$ the functional: $a \mapsto \langle \pi(a)\xi, \eta \rangle$ is weak-star continuous. Also, by norm-boundedness it suffices to prove so for a norm-dense subset of such $(\xi, \eta)$. If $\xi$ and $\eta$ are the images of elements $x, y \in A$, then this functional is equal to $\phi(y^*ax)$, so it is continuous. □

Now we form the representation $\mathcal{H}_n = \oplus_{\phi \in \mathscr{S}_n} \mathcal{H}_\phi$. It follows that the corresponding map $\pi_n : A \to B(\mathcal{H}_n)$ is weak-star continuous (again, by the fact that it is norm-bounded and the functionals $a \mapsto \langle \pi_n(a)\xi, \eta \rangle$ are continuous for a norm-dense subset of $(\xi, \eta)$). By Proposition **??**, it is injective, and by Theorem **??** we get:

**Theorem 6.3.** *Every $W^*$-algebra is isomorphic to a weak-star closed *-subalgebra of $B(\mathcal{H})$, for some Hilbert space $\mathcal{H}$, and vice versa.*

Finally, a *-subalgebra of $B(\mathcal{H})$ is weak-star closed if and only if it is closed in the weak operator topology, so we might as well replace the former by the latter in the previous theorem.

## 20.7 Projections in a $W^*$-algebra

`Later`

# Part III

# Algebraic groups and their automorphic quotients

# Chapter 21

# Basic notions

`This chapter is a repetition of material presented in Part ??.`

## 21.1 References

- A. Borel, *Linear Algebraic Groups.* (2nd ed., Springer).
- T. A. Springer, *Linear Algebraic Groups.* (2nd ed., Birkhäuser).
- T. A. Springer, *Reductive Groups.* Automorphic forms, representations and $L$-functions, PSPUM 33.1, AMS (the "Corvallis proceedings").

## 21.2 Basic notions

Let $k$ be a field.

A *group scheme* is a group in the category of $k$-schemes. Facts: If $k$ is of characteristic zero, then any group scheme is *smooth* over $k$. In characteristic $p > 0$, though, a group scheme could be non-reduced, for instance the *Frobenius* group scheme, i.e. the kernel of $\mathbb{G}_m \to \mathbb{G}_m : g \mapsto g^p$. An *algebraic group* is a smooth group scheme over a field $k$. FROM NOW ON, ONLY CHARACTERISTIC ZERO.

The geometrically connected component $G^0$ of $G$ is always defined over $k$ and normal. (The group of components $G/G^0$ is also an algebraic group over $k$, cf. "homogeneous spaces".)

The *multiplicative group* $\mathbb{G}_m = \operatorname{spec} k[T, T^{-1}], T \mapsto T_1 \cdot T_2$. The *additive group* $\mathbb{G}_a = \operatorname{spec} k[T], T \mapsto T_1 + T_2$.

A *linear algebraic group* is one which is affine. Equivalently, it admits a closed embedding into $\mathrm{GL}_n$ for some $n$.

The *Jordan decomposition*: Every element $g \in \mathrm{GL}_n$ can be written uniquely as $g_s g_u$, with $g_s$ *semisimple* and $g_u$ *unipotent*, so that $g_s$ and $g_u$ commute. If $g \in \mathrm{GL}_n(k)$, then so are $g_s$ and $g_u$. Let $G$ be a linear algebraic group and

$i : G \hookrightarrow \mathrm{GL}_n$ an embedding. Let $g \in G$. Facts: whether $i(g)$ is semisimple or unipotent does not depend on the embedding! Therefore $g$ will be called semisimple or unipotent accordingly. Moreover, $g = g_s g_u$ uniquely, with $g_s \in G$ semisimple and $g_u \in G$ unipotent.

From now on "group" will mean "linear algebraic group" (over $k$).

## 21.3 Homogeneous spaces

If $G$ is a group and $H$ a closed subgroup, then $H\backslash G$, the *geometric quotient* of $G$ by $H$, exists in the category of $k$-schemes, and is also smooth over $k$. (The existence statement is valid in positive characteristic as well, but of course the quotient may not be reduced if $G$ is not.) However, *it quasi-projective but not necessarily affine.*

(Definition of geometric quotient $\pi : X \to Y$ of a scheme $X$ by the action of $G$: $\pi$ is open and surjective, $\mathcal{O}_Y = (\pi_* \mathcal{O}_X)^G$ and the geometric fibers are precisely the geometric orbits of $G$. For more details see, for instance, Mumford, *Geometric Invariant Theory* or the fourth chapter in the book of van der Geer and Moonen on abelian varieties, appearing in preliminary form at: `http://staff.science.uva.nl/~bmoonen/boek/BookAV.html`.)

The idea of the proof: Show that there exists an algebraic representation $V$ of $G$ such that $H$ is the stabilizer of a line in $V$. One finds such a representation by playing with the ring of regular functions on $G$ and the ideal defining $H$. Then $H\backslash G$ can naturally be identified with the orbit of that line in $\mathbb{P}(V)$. (General facts about algebraic representations and actions of $G$ on varieties: By definition, their matrix coefficients are regular functions on $G$. Every action of $G$ on an affine variety $X$ is locally finite, i.e. every $f \in k[X]$ generates a finite-dimensional subspace. Every orbit of $G$ on $X$ is locally closed.)

Example : $P\backslash G$, where $P$ is a parabolic subgroup, is projective. (This is actually the definition of *parabolic subgroup*, so the statement is void unless you know some examples of parabolic subgroups!) In fact, for reductive $G$ the quotient $H\backslash G$ is affine if and only if $H$ is also reductive. (Borel & Harish-Chandra.) We will explain these notions later.

If $H$ is normal, then $H\backslash G$ has a natural group structure. Beware: the points of $(H\backslash G)(k)$ are not the same as $H(k)\backslash G(k)$, in general. Example: the space $O_n \backslash \mathrm{GL}_n$ of non-degenerate quadratic forms in $n$-variables.

## 21.4 Diagonalizable groups

The character group $\mathcal{X}(G)$ of $G$ is the group of morphisms: $G \to \mathbb{G}_m$. If $F$ is a superfield of the field of definition, we will write: $\mathcal{X}_F(G)$ for $\mathcal{X}(G_F)$ (where $G_F$ denotes the base change to $F$), i.e. those characters defined over $F$.

*Exercise.* The only character of $\mathbb{G}_a$ is the trivial one.

(This has the following generalization: The only irreducible algebraic representation of a unipotent group is the trivial one.)

On the other hand, the characters of $\mathbb{G}_m$ are precisely those of the form $T \mapsto T^r$ for some $r \in \mathbb{Z}$. We notice that if we consider them as functions $\mathbb{G}_m \mapsto \mathbb{G}_m \hookrightarrow \mathbb{A}^1$ then *every element of* $k[\mathbb{G}_m]$ *can be written as a linear combination of characters.* (In other words, $k[\mathbb{G}_m]$ is the Grothendieck ring over $k$ of the group $\mathcal{X}_{\bar{k}}(\mathbb{G}_m)$.)

A *diagonalizable group* is one for which its $\bar{k}$-character group $\mathcal{X}_{\bar{k}}(G)$ spans $\bar{k}[G]$ over $\bar{k}$. Equivalently, it is commutative and semisimple. It is called *split* according as $\mathcal{X}_k(G) = \mathcal{X}_{\bar{k}}(G)$ or not. A *torus* is a group which is isomorphic, over $\bar{k}$, to $\mathbb{G}_m^r$ for some $r$. The connected component of a smooth diagonalizable group is a torus.

**Theorem 4.1.** *The contravariant functor $G \mapsto \mathcal{X}_{\bar{k}}(G)$ is an equivalence between the categories of diagonalizable $k$-groups and of finitely generated $\mathbb{Z}$-modules with a (continuous)* $\mathrm{Gal}(\bar{k}/k)$*-action.*

This is a relief: Everything that we would like to know about these groups can be reduced to a relatively simple combinatorial picture. Let us see an example:

*Example* 4.2. The character group of $\mathbb{G}_m^2$ is isomorphic to $\mathbb{Z}^2$; fix such an isomorphism. Let $E$ be a quadratic extension of $k$. There is a diagonalizable group $R$ which is isomorphic to $\mathbb{G}_m^2$ over the algebraic closure, but such that $R(k) = E^\times$.[1] Let $\sigma$ be the non-trivial element of $\mathrm{Gal}(E/k)$. Then we define the action of $\mathrm{Gal}(\bar{k}/k)$ on $\mathcal{X}(\mathbb{G}_m^2)$ as follows: It will factor through $\mathrm{Gal}(E/k)$, and ${}^\sigma(a,b) = (b,a) \in \mathbb{Z}^2$. The short exact sequence:

$$0 \to \{(a,-a) | a \in \mathbb{Z}\} \to \mathbb{Z}^2 \to \mathbb{Z} \to 0$$

(where $\mathrm{Gal}(\bar{k}/k)$ acts trivially on the last group) corresponds, dually, to a sequence of the groups:

$$1 \leftarrow R/\mathbb{G}_m \leftarrow R \leftarrow \mathbb{G}_m \leftarrow 1$$

which at the level of points is the embedding $k^\times \hookrightarrow E^\times$.

On the other hand there is another short exact sequence:

$$0 \to \{(a,a) | a \in \mathbb{Z}\} \to \mathbb{Z}^2 \to \mathbb{Z} \to 0$$

where now the Galois group acts trivially on the first term, but not on the last. This corresponds to the norm map:

$$E^\times \to k^\times.$$

Notice that the kernel of the norm map is an 1-dimensional torus. Moreover, notice that here we have a surjection of algebraic groups which is not a surjection at the level of $k$-points.

[1] We recall here the notion of *points* of a scheme: Given schemes $X, Y$ over a basis $S$, we define $X(Y) := \mathrm{Mor}(Y, X)$ in the category of $S$-schemes (the "$Y$-points of $X$"). When $Y = \operatorname{spec} A$, for a ring $A$, we also say "$A$-points". If $S = \operatorname{spec} k$, $X$ is affine and $A$ is any $k$-algebra then $X(A) = \mathrm{Hom}(k[X], A)$. Of course, for our discussion, $G(A)$ is the same as $i(G) \cap \mathrm{GL}_n(A)$ for any closed embedding $i : G \to \mathrm{GL}_n$.

*Remark.* The group constructed above is called the *restriction of scalars* of $\mathbb{G}_m$ from $E$ to $k$; namely, we start with the group $\mathbb{G}_m$ over $E$ and construct an algebraic group $R$ over $k$ such that for every extension $F$ of $k$ we have $R(F) = \mathbb{G}_m(F \otimes E)$. We will return to this concept in a more general setting.

*Exercise.* Verify the above-mentioned property of restriction of scalars for the given example.

*Exercise.* Prove the following: If $k$ is algebraically closed and $D$ is a diagonalizable group, then there exists a finite subgroup $F$ of $D$ such that $D = D^0 \times F$. Does that hold for arbitrary $k$?

## 21.5 Reductive groups

A group is *solvable* if it admits a normal series whose successive quotients are abelian. Notice that this notion is $k$-independent, since we can take the normal series where $G_i$ is followed by $[G_i, G_i]$, which is defined over $k$. A *Borel subgroup* of a linear group $G$ is one which is maximal (over the algebraic closure) among the connected solvable subgroups. It is not true in general that there is a Borel subgroup over $k$; if there is, the group $G$ will be called *quasi-split*. We will see later that every reductive group is an *inner form* of precisely one (isomorphism class of) quasi-split group.

Facts: All Borel subgroups are $G(k)$-conjugate. A subgroup is parabolic if and only if it contains a Borel. (The proof of these facts relies on (a generalization of) the Lie-Kolchin Theorem, which states that a connected solvable group acting on a proper variety has a fixed point.)

Every group $G$ has a maximal closed, connected, normal, solvable subgroup. This is the *radical* $\mathcal{R}(G)$. The unipotent elements of $\mathcal{R}(G)$ form a maximal closed, connected, unipotent subgroup of $G$, the *unipotent radical* $\mathcal{R}_u(G)$. We say that $G$ is *reductive* if $\mathcal{R}_u(G) = 1$ and *semisimple* if $\mathcal{R}(G) = 1$. The *rank* of $G$ is the dimension of a maximal torus and its *semisimple rank* is the rank of its derived group; equivalently, the dimension of its root system (see below).

(Discussion of examples in class.)

*Reductivity theorem:* If $G$ is reductive then every algebraic representation of $G$ is completely reducible. This theorem does not hold in positive characteristic, where it is replaced by *geometric reductivity* (Haboush). (The latter states that for every $G$-invariant vector there is a $G$-invariant function, homogeneous of degree a power of the characteristic exponent of the field, which is non-zero on the given vector. If the characteristic exponent is 1, this leads to complete reducibility.)

The notion of semisimplicity is better understood at the level of Lie algebras. *Lie algebra* $\mathfrak{g}$ of $G$ = tangent space at the identity = derivations of $k[G]$ in $k[G]$ invariant by left translation = differential operators on $G$ homogeneous of degree 1 invariant by left translation. All the definitions given above (solvable, reductive, semisimple etc.) have obvious analogs for Lie algebras and a (connected) group $G$ is ... if its Lie algebra is ... .

A Lie algebra $\mathfrak{g}$ is called *simple* if it is not abelian and its only ideals are 0 and $\mathfrak{g}$. Then: $\mathfrak{g}$ is semisimple if and only if it is isomorphic to a product of simple algebras. This is equivalent to There is also another criterion for when a Lie algebra is semisimple, sometimes taken as the definition: $\mathfrak{g}$ is semisimple iff the invariant bilinear form: $(X, Y) \mapsto \mathrm{tr}(\mathrm{ad}(X) \circ \mathrm{ad}(Y))$ (the *Killing form*) is non-degenerate. A very nice reference for semisimple Lie algebras is the book of Serre: *Complex Semisimple Lie Algebras*.

## 21.6 Root systems and root data

For this section we assume $k$ to be algebraically closed. In the next lecture we will discuss rationality issues.

We saw a nice combinatorial description for diagonalizable groups, we would have liked to have the same for more general classes of groups. This is not possible in the sense of getting an equivalence of categories[2], but at least we can fully describe the isomorphism classes of groups this way.

More precisely, the *adjoint representation* $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ of a reductive group gives rise to its root datum: This is a combinatorial description of the structure of $G$.

Let $G$ be reductive, and let $A$ be a maximal torus in $G$. We denote by small gothic letters the corresponding Lie algebras. The group $A$ is equal to its centralizer; the centralizer of a maximal torus in a (not necessarily reductive) algebraic group is called a *Cartan subgroup*. All Cartan subgroups are conjugate. Under the adjoint representation, $\mathfrak{g}$ decomposes into eigenspaces for $A$:

$$\mathfrak{g} = \mathfrak{a} \oplus \sum_{\alpha \in \Phi} \mathfrak{u}_\alpha.$$

Facts: $\mathfrak{a}$ is precisely equal to the zero eigenspace (i.e. $\mathfrak{a}$ is its own commutator). The non-zero eigencharacters of $A$ are called *roots* (their set denoted by $\Phi$), and their eigenspaces $\mathfrak{u}_\alpha$ are all one-dimensional.

Recall that we study diagonalizable groups via their character lattices. The character lattice $\mathcal{X}(A)$ is free, and $\Phi$ is a finite subset of it. The *Weyl group* $W := \mathcal{N}(A)/A$ acts on $\mathcal{X}(A)$ and the set of roots. It is a finite group.

Let us first consider the case of $\mathrm{SL}_2$: Its Lie algebra is generated by three elements $h = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$, $e = \begin{pmatrix} & 1 \\ & \end{pmatrix}$ and $f = \begin{pmatrix} & \\ 1 & \end{pmatrix}$ which satisfy the commutation relations $[h, e] = 2e, [h, f] = -2f$ and $[e, f] = h$. Three non-zero elements $(h, e, f)$ in a Lie algebra which satisfy these commutation relations are called an $\mathfrak{sl}_2$*-triple*. Here we have $\mathcal{X}(A) \simeq \mathbb{Z}$, with roots $\pm\alpha = \pm 2$ under such an isomorphism and $W = \mathbb{Z}/2$ which sends $\alpha$ to $-\alpha$.

[2]There is a good reason for it: To get an equivalence of categories one must consider the category of all $G$-representations, cf. Tannaka-Krein duality. For diagonalizable groups this category is described easily in terms of combinatorial data, this is no longer the case for other groups.

Similarly, for every reductive group if $\alpha \in \Phi$ then $-\alpha \in \Phi$ and there are $h \in \mathfrak{a}$, $e \in \mathfrak{u}_\alpha$ and $f \in \mathfrak{u}_{-\alpha}$ such that $(h, e, f)$ is an $\mathfrak{sl}_2$-triple in $\mathfrak{g}$. In fact, $h$ is unique, $e$ can be chosen to be an arbitrary non-zero element and $f$ is unique once $e$ is chosen. The element $h$ lives in $\mathcal{X}(A)^* := \mathrm{Hom}_{\mathbb{Z}}(\mathcal{X}(A), \mathbb{Z}) = \mathrm{Hom}(\mathbb{G}_m, A)$, and it is called the *co-root* $\check{\alpha}$ associated to $\alpha$. Hence, to the group $G$ one can also associate the set $\check{\Phi}$ of co-roots, a finite subset of $\mathcal{X}(A)^*$.

### 21.6.1 Root systems

Both the sets of roots and of co-roots, together with the action of the Weyl group, each satisfy the axioms for a combinatorial object called *the root system*. Strictly speaking, a root system is formed by $\Phi$ and its linear span $V$ in $\mathcal{X}(A) \otimes \mathbb{R}$ (and similarly for the co-roots). These axioms are:

1. $\Phi$ is a finite subset spanning $V$ and not containing 0.
2. For every $\alpha \in \Phi$ there is a symmetry $w_\alpha$ (that is, a linear automorphism of $V$ whose fixed point set is a hyperplane) such that $w_\alpha(\Phi) = \Phi$ and $w_\alpha(\alpha) = -\alpha$.
3. For each $\alpha, \beta \in \Phi$, $w_\alpha(\beta) - \beta$ is an integer multiple of $\alpha$.

Moreover, the ("absolute") root system arising from a reductive group is *reduced*, or *crystallographic*: If $\alpha, c\alpha \in \Phi$ then $c = \pm 1$.

Now the reader should consult a reference on root systems (for instance, Serre's book) to learn more about them, their classification, etc. In short: Since $W$ is a finite group, it is possible to find an inner product on $V$ such that the $w_\alpha$'s are reflections. This inner product is unique up to multiple on every irreducible component (a root system is irreducible if it is not the sum of two sub-root systems). With respect to this inner product, for every $\alpha, \beta \in \Phi$ we have $w_\alpha(\beta) = \beta - 2\frac{(\alpha,\beta)}{(\alpha,\alpha)}\alpha$. The angle between $\alpha$ and $\beta$, for $\alpha$ and $\beta$ simple roots (for a definition of simple roots, see the next lecture), encodes the way in which $w_\alpha$ and $w_\beta$ commute (more precisely, $W$ is the group with generators $w_\alpha$, with $\alpha$ ranging over simple roots, and relations $w_\alpha^2 = 1$ and $(w_\alpha w_\beta)^{m(\alpha,\beta)} = 1$ where $m(\alpha, \beta) = 2, 3, 4$ or 6 according as the angle between *alpha* and $\beta$ is $\pi/2, 2\pi/3, 3\pi/4,$ or $5\pi/6$). It also encodes relations for a set of generators of the Lie algebra $\mathfrak{g}$. The simple roots are the vertices of the *Dynkin diagram*, with an edge between two roots iff they are not orthogonal – the edge being simple, double or triple according as the angle between them is $2\pi/3, 3\pi/4,$ or $5\pi/6$, and an arrow from the longest to the shortest, if they don't have the same length. The reduced, irreducible root systems have been classified: there are four infinite families $A_n, B_n, C_n, D_n$ and five *exceptional root systems* $G_2, F_4, E_6, E_7, E_8$.

(Examples in class.)

*Remark.* The (hidden) condition that the Weyl group be finite imposes strong combinatorial restrictions that lead to only this finite number of families. There is a more general world, set up in a way that does not include the condition of finiteness, that of *Kac-Moody algebras*.

### 21.6.2 Root data

Notice that the root system only determines the adjoint group of $G$, for instance the groups $\mathrm{GL}_n$, $\mathrm{SL}_n$, $\mathrm{PGL}_n$ all have the same root system. We must find a way to encode information about the center of the group.

The data $(X := \mathcal{X}(A), \Phi, \check{X} := \mathcal{X}(A)^*, \check{\Phi})$ satisfy the axioms for a *root datum*. These consist of:

1. Two lattices (free $\mathbb{Z}$-modules of finite type) which are each other's dual.
2. Finite subsets $\Phi \subset X, \check{\Phi} \subset \check{X}$ with a bijection $\Phi \to \check{\Phi} : \alpha \mapsto \check{\alpha}$ such that $\langle \alpha, \check{\alpha} \rangle = 2$ (to understand this condition, recall the case of $\mathfrak{sl}_2$).
3. The endomorphisms of $X, \check{X}$ defined by $w_\alpha(x) := x - \langle x, \check{\alpha} \rangle \, \alpha$, $w_{\check{\alpha}}(\check{x}) = \check{x} - \langle \alpha, \check{x} \rangle \, \check{\alpha}$ preserve $\Phi$ and $\check{\Phi}$.

(The last axiom is equivalent to: The endomorphisms $w_\alpha$ preserve $\Phi$ and generate a finite group.)

**Theorem 6.1** (Classification over the algebraic closure)**.** *Assume $k$ algebraically closed. For any root datum there exists a connected reductive group $G$ and a maximal torus $A \subset G$ over $k$ which give rise to that root datum; the pair $(G, A)$ is unique up to isomorphism.*

Some more definitions: Let $\mathcal{R} \subset X, \check{\mathcal{R}} \subset \check{X}$ be the lattices spanned by roots, resp. co-roots, $V = X \otimes \mathbb{R}$ and $\check{V} = \check{X} \otimes \mathbb{R}$. The root datum is called *semisimple* if $\mathcal{R} \otimes \mathbb{R} = V$; this is equivalent to the corresponding group being semisimple. Assume that this is the case. Let $\mathcal{P} \subset V$, $\check{\mathcal{P}} \subset \check{V}$ be the duals of $\check{\mathcal{R}}, \mathcal{R}$ respectively. Hence $\mathcal{R} \subset X \subset \mathcal{P}$ and $\check{\mathcal{R}} \subset \check{X} \subset \check{\mathcal{P}}$ We say that the root datum is *simply-connected* if $X = \mathcal{P}$ and *adjoint* if $X = \mathcal{R}$. This is known to be equivalent to the corresponding group being so. (Simply connected means that the étale fundamental group is trivial; in the complex case this is equivalent to the topological fundamental group being trivial; adjoint means that $G = G^{\mathrm{ad}}$.)

Examples: $\mathrm{SL}_n$ is simply connected, $\mathrm{PGL}_n$ is adjoint.

The elements of $\mathcal{P}$ (resp. $\check{\mathcal{P}}$) are called *weights* (resp. co-weights). Their significance is the following: Notice first that they depend only on the root system, not the root datum. Hence, they can be "seen" by the Lie algebra, which does not "see" the precise character lattice of the maximal torus. Over an algebraically closed field, dominant weights parametrize isomorphism classes of irreducible finite-dimensional modules for the Lie algebra. Then these representations can be "lifted" to the group (i.e. are the differential of a group representation) if and only if the corresponding weight is *integral*, which means that it belongs to $X$. For instance, if the group is simply connected then all representations can be lifted. (No surprise here!)

While it is difficult to describe morphisms between algebraic groups in general, it is easy to do so for morphisms which are *central isogenies*; that is, surjective morphisms with finite kernels belonging to the center. The combinatorial shadow of a central isogeny is the notion of *isogeny of root data*: Given

two sets of root data $\Psi = (X, \Phi, \check{X}, \check{\Phi})$ and $\Psi' = (X', \Phi', \check{X}', \check{\Phi}')$ a homomorphism $X \to X'$ is called an isogeny if it is injective with image of finite index in $X'$, it maps $\Phi$ bijectively to $\Phi'$ and its adjoing maps $\check{\Phi}'$ bijectively to $\check{\Phi}$.

**Theorem 6.2.** *Assume $k$ algebraically closed. Consider root data as above and let $(G, A)$, $(G', A')$ be the corresponding groups and maximal tori according to* **??**. *Every central isogeny $(G, A) \to (G', A')$ induces canonically an isogeny of root data $\Psi' \to \Psi$. Conversely, if $f : \Psi' \to \Psi$ is an isogeny, there exists a central isogeny $G \to G'$, unique up to automorphisms* $\mathrm{Inn}(a), a \in A$.

This theorem will allow us to understand automorphisms of reductive groups, which in turn will help us understand their forms over a non-algebraically closed field. First, we need to discuss based root data:

### 21.6.3 Weyl chambers and based root data

Given a root system $(V, \Phi)$ with Weyl group $W$, the complement $\mathring{V}$ of the union of fixed point hyperplanes of reflections in $W$ is the union of a finite number of connected components, each of which is a fundamental domain for the action of $W$ on $\mathring{V}$. Choosing such a *Weyl chamber* $\mathcal{C}$ gives rise to two translation-invariant orderings on $V$: The stronger of the two is the one defined by $C$, and elements which are greater or equal to zero with respect to this (i.e. elements of $\overline{\mathcal{C}}$) are called *dominant*. The weaker is the one defined by the cone spanned by the roots $\alpha$ such that $\langle c, \check{\alpha} \rangle > 0$ for every $c \in \mathcal{C}$. These roots are called *positive* and their set will be denoted by $\Phi^+$. Notice that $\forall \alpha \in \Phi, \alpha \in \Phi^+$ or $-\alpha \in \Phi^+$. When expressing an inequality in $V$ we will, by default, refer to this weaker ordering; when we say "an ordering of the roots" or "a choice of positive roots" we mean one induced by the choice of a Weyl chamber. The irreducible elements of $\Phi^+$ (i.e. those which cannot be written as a sum of others) form the set of *simple roots*, usually denoted by $\Delta$, and they are a basis for $V$. A choice of positive roots induces a choice of positive co-roots in the dual root system, such that $\alpha > 0 \iff \check{\alpha} > 0$.

If $(G, A)$ is an algebraic group with a maximal torus (over an algebraically closed field), $(V, \Phi)$ is the corresponding root system and $B$ is a Borel subgroup containing $A$, then it is easy to see that the roots of $\mathfrak{b}$ form the set of positive roots with respect to a choice of Weyl chamber. One can see that the set of Weyl chambers is in bijection with the Borel subgroups of $G$ containing $A$. A root datum $\Psi$, together with a choice of positive roots as above, is called a *based root datum*. An automorphism of a based root datum is an automorphism of the root datum which preserves the set of positive roots.

### 21.6.4 Automorphisms

Now we study automorphisms of a reductive group $G$. Fix a maximal torus $A$, a Borel $B \supset A$ (corresponding to a choice of positive roots) and non-zero elements $u_\alpha \in \mathfrak{u}_\alpha$ for all $\alpha > 0$. Given an automorphism $\phi$ of $G$, by the fact that all Borels are conjugate we can compose it with an inner automorphism to get

an automorphism $\phi'$ which preserves $B$. Moreover we can compose that with $\mathrm{Inn}(a)$, for some $a \in A$, to get an automorphism $\phi''$ which fixes the $u_\alpha$'s. It is a fact that the Borel subgroup is its own normalizer; from this it can be seen that $\phi''$ is the only element in its Inn-coset which preserves the data $(A, B, \{u_\alpha\}_\alpha)$.

Hence:

$$1 \to \mathrm{Inn}(G) \to \mathrm{Aut}(G) \to \mathrm{Out}(G) = \mathrm{Aut}(G, A, B, \{u_\alpha\}_\alpha) \to 1$$

and this sequence splits.

It is clear from our discussion of central isogenies that automorphisms of $(G, A, B, \{u_\alpha\}_\alpha)$ are in bijection with automorphisms of the corresponding based root datum. (For a semisimple simply connected or adjoint group the latter are, in turn, the same as *isomorphisms of the Dynkin diagram.*)

In the next lecture, we will use our knowledge of automorphisms to discuss the Galois action on a group and classify forms of the group over a non-algebraically closed field.

## 21.7 Parabolic subgroups

**Theorem 7.1.** *Let $G$ be a reductive group and fix a based root datum corresponding to $(A, B)$. Let $\Delta$ denote the set of simple positive roots. The parabolic subgroups of $G$ containing $B$ are in natural bijection with subsets of $\Delta$. Let $I \subset \Delta$ and let $P_I$ be the corresponding subgroup, $U_I$ its unipotent radical. Then $P_I = L_I \ltimes U_I$ where $L_I$ is the following subgroup: We have $\mathcal{Z}(L_I) = \{a \in A | \alpha(a) = 0$ for all $\alpha \in I$; and $L_I$ is the centralizer in $G$ of $\mathcal{Z}(L_I)$.*

Such a subgroup $L_I$ is called a *Levi subgroup*. The equalities above also hold at the level of $k$-points for any field $k$, if $P_I$ is defined over $k$: $P_I(k) = L_I(k)U_I(k)$. Moreover, $(G/P)(k) = G(k)/P(k)$ for every parabolic. Having fixed a Borel subgroup $B$, a parabolic which contains it is called a *standard parabolic*. Since all Borels are conjugate, all parabolics are conjugate to a (unique) standard one.

# Chapter 22

# Structure and forms over a non-algebraically closed field.

## 22.1 Restriction of scalars

We saw in the previous lecture that for an extension $E/k$ we can regard $E^\times$ either as $\mathbb{G}_m(E)$ or as the $k$-points of a group over $k$. One is of course familiar with such a situation: the additive group $\mathbb{C}$ is isomorphic to $\mathbb{R}^2$. We will now discuss this in more generality.

Let $E/k$ be an extension, $X$ a variety over $E$. We would like to define a variety $Y$ over $k$ such that the $k$-points of $Y$ are naturally identified with the $E$-points of $X$. In fact, we would like this to hold also for the $S$-points of $Y$, where $S$ is any scheme over $k$ (for instance, the $F$-points, where $F$ is an extension of $k$); those should be the same as $S \times_{\operatorname{spec} k} \operatorname{spec} E$ points of $X$.

**Definition.** Consider the functor {Schemes over k} $\to$ {Sets} defined by $S \mapsto X(S \times_{\operatorname{spec} k} \operatorname{spec} E)$. If this functor is represented by a scheme $Y$ over $k$, we call $Y$ the *restriction of scalars* of $X$ from $E$ to $k$, denoted $\operatorname{Res}_{E/k} X$.

The definition says nothing more than the preceeding paragraph. But from the wording of the definition + abstract nonsense it follows that if the $k$-scheme $\operatorname{Res}_{E/k} X$ exists then it is unique up to unique isomorphism.

The following holds in arbitrary characteristic:

**Theorem 1.1.** *If $X$ is an irreducible non-singular affine variety over $E$ then $\operatorname{Res}_{E/k} X$ exists and is also irreducible, non-singular and affine. If $E/k$ is separable, the same holds without the denominations "irreducible" and "non-singular". If $E/k$ is separable and $X$ is a reductive group then $\operatorname{Res}_{E/k} X$ is also reductive.*

*Example* 1.2. Let $E/k$ be a quadratic Galois extension, fix a non-degenerate hermitian form on $E^n$ and let $U \subset GL_n(E)$ denote the corresponding unitary group. Then $U$ cannot be identified with the group of points of an $E$-subgroup of $GL_n$; however, $U = U_n(k)$, where $U_n$ is a $k$-subgroup of $\mathrm{Res}_{E/k}\,\mathrm{GL}_n$.

*Remark* (Important!). Every group defined over a number field can, by restriction of scalars, be treated as a group over $\mathbb{Q}$. This turns out to be very convenient in some general applications, such as the results of Borel and Harish-Chandra that we will discuss soon. However, by restriction of scalars we lose some good properties, such as that of being split. (Being quasi-split, however, is a property which is preserved.)

## 22.2 Structure

Let $G$ be a reductive group over a field $k$ (in characteristic zero).

Cartan subgroups are always defined over $k$; Borel subgroups are not. If there exist Borel subgroups over $k$ the group is called *quasi-split*. If those contain a split maximal torus (equivalently: if there exists a split maximal torus) then the group is called *split*. But in general, we fix a Cartan subgroup $A$ and a maximal split torus $S \subset A$. The rank of $S$ is the *$k$-rank* of $G$. We let again $S$ act on the Lie algebra of $G$, and mark the non-zero eigenvalues on $\mathcal{X}(S)$; these give rise to the *relative root system* $\Phi_k$ of $G$. Notice that in this case the root subspaces are not necessarily one-dimensional; and under the restriction map $\mathcal{X}_{\bar{k}}(A) \to \mathcal{X}(S)$ some (absolute) roots of $G$ may map to zero. More precisely:

**Theorem 2.1.** *The group $\mathcal{Z}(S)$ is the Levi subgroup of a minimal parabolic $k$-subgroup of $G$. Given an ordering of $\Phi_k$ with set of simple roots $\Delta_k$ there is a compatible ordering of $\Phi$ (with simple roots $\Delta$) (i.e. such that $\Delta_k$ is the image of those elements of $\Delta$ which don't map to zero). There is a bijection between standard parabolic $k$-subgroups of $G$ and subsets of $\Delta_k$. All parabolic $k$-subgroups are conjugate under $G(k)$ to a (unique) standard one.*

A group is called *anisotropic* if $S$ is trivial, *isotropic* otherwise.

## 22.3 Forms

**Definition.** Let $X$ be a $k$-scheme. A *form* (or *twist*) of $X$ is a $k$-scheme $X'$ such that $X'_{\bar{k}}$ is isomorphic to $X_{\bar{k}^s}$.

From now on, let us work for simplicity in the category of quasi-projective $k$-varieties. Let $\Gamma = \mathrm{Gal}(\bar{k}/k)$.[1] We plan to explain the following theorem:

**Theorem 3.1.** *Isomorphism classes of forms of $X$ are in natural bijection with $H^1(\Gamma, \mathrm{Aut}_{\bar{k}}(X))$.*

[1] We gave the general definition using the separable closure of $k$ because everything that we say in this section holds in arbitrary characteristic. In any case, if $X = G$ an algebraic group, any $X'$ which is $\bar{k}$-isomorphic to $G$ is also $\bar{k}^s$-isomorphic.

Here $H^1$ is the cohomology group defined by *continuous* cocycles from $\Gamma$ to $\mathrm{Aut}_{\bar{k}}(X)$, the latter considered discrete. From now on we will write simply $H^*(A)$ for the Galois cohomology of a $\Gamma$-group $A$.

*Remark.* The natural isomorphism of the theorem depends on the form $X$ that we started with, as does the action of $\Gamma$ on $\mathrm{Aut}_{\bar{k}}$. More precisely, as we will recall, $H^1$ is a pointed set, and the distinguished point corresponds to the class of $X$.

Let $X'$ be a form of $X$, and $\phi : X_{\bar{k}} \to X'_{\bar{k}}$ an isomorphism. For any $\gamma \in \Gamma$ the map $\phi^{-1} \circ \gamma \circ \phi \circ \gamma^{-1}$ is an automorphism of $X_{\bar{k}}$. Define:

$$c : \Gamma \to \mathrm{Aut}(X)$$

by $\gamma \mapsto \phi^{-1} \circ \gamma \circ \phi \circ \gamma^{-1}$. It satisfies the *cocycle* condition:

$$c(\gamma_1\gamma_2) = c(\gamma_1) \cdot {}^{\gamma_1}c(\gamma_2).$$

Conversely, given a (continuous) cocycle define the following action $*$ of $\Gamma$ on $X(\bar{k})$: $\gamma * x := c(\gamma)({}^{\gamma}x)$ (where the exponent, as usual, denotes the given action). It can be seen that the new action comes from a $k$-form of $X$.

On the set of cocycles from $\Gamma$ to a $\Gamma$-group $A$ we define the equivalence relation: $c \sim c'$ iff $c(\gamma) = b^{-1}c'(\gamma){}^{\gamma}b$ for some $b \in \mathrm{Aut}_{\bar{k}}(X)$. The set of equivalence classes is $H^1(\Gamma, A)$. It is not a group if $A$ is not abelian; however, it has a distinguished point coming from the trivial cocycle (whose equivalence class is the set of *coboundaries*).

Non-abelian cohomology groups have the usual "long exact sequence" property of derived-functor cohomology for the functor $A \mapsto H^0(A) := A^{\Gamma}$, as long as the long exact sequence makes sense. That is, if we have a short exact sequence of $\Gamma$-groups:

$$1 \to A \to B \to C \to 1$$

then we get

$$1 \to H^0(A) \to H^0(B) \to H^0(C) \to H^1(A) \to H^1(B) \to H^1(C)$$

as maps of pointed sets (i.e. the distinguished point maps to the distinguished point). "Long exact" means that the preimage of the distinguished point is the image of the previous arrow. Moreover, if $A$ is central in $B$, then this sequence extends to include $H^2(A)$.

The non-existence of a group structure on $H^1$ may cause some confusion. For instance: The $\Gamma$-structure on $\mathrm{Aut}(X)$ depends on the form $X$ chosen, but the set of isomorphism classes of forms of $X$ does not. To explain this, one can check that if $X'$ is a form, $\phi : X \to X'$ and $c$ the corresponding cocycle, then for every 1-cocycle $c*$ with values in $\mathrm{Aut}(X')$ the map $\phi^{-1}(c*) \cdot c$ is a 1-cocycle in $\mathrm{Aut}(X)$ and this establishes an isomorphism of sets (*not* pointed sets, as 1 goes to $[c]$!): $H^1(\mathrm{Aut}(X')) \to H^1(\mathrm{Aut}(X))$.

## 22.4 Forms of reductive groups

Let $G$ be a quasi-split reductive group over $k$, $A \subset B$ a maximal torus and a Borel *over* $\bar{k}$.

In the previous lecture we established the split short exact sequence:

$$1 \to \mathrm{Inn}(G) \to \mathrm{Aut}(G) \to \mathrm{Out}(G) = \mathrm{Aut}(\Psi(G, A, B)) \to 1 \tag{22.1}$$

(where we mean $\bar{k}$-automorphisms but we drop the subscripts $_{\bar{k}}$ for simplicity; $\Psi(G, A, B)$ denotes the corresponding based root datum).

The $k$-structure on $G$ defines an action of $\Gamma$ on $\mathrm{Aut}(X)$ which preserves $\mathrm{Inn}(G)$. More precisely, we have a canonical $\Gamma$-isomorphism: $\mathrm{Inn}(G) = G^{\mathrm{ad}}$ where $G^{\mathrm{ad}} = G/\mathcal{Z}(G)$ denotes the adjoint group of $G$. The splitting can also be chosen so that it is preserved by $\Gamma$.

Hence we have: $H^1(\mathrm{Inn}(G)) \to H^1(\mathrm{Aut}(G))$. Forms of $G$ whose isomorphism classes correspond to elements in the image of $H^1(\mathrm{Inn}(G))$ are called[2] *inner forms*. The discussion at the end of the previous subsection shows that $G_1$ being an inner form of $G_2$ is an equivalence relation.

From the splitting of (**??**) one can deduce that: Every isomorphism class is an inner form of *precisely one quasi-split form*. Assume now that $G$ is quasi-split; then (**??**) splits in a $\Gamma$-equivariant way.

**Definition.** A group $G$ over $k$ is called *unramified* if it is quasi-split, and splits over an unramified extension.

**Theorem 4.1.** *Let $k$ be a global field.*[3] *Let $G$ be a connected reductive group over $k$, $G_v := G \times \operatorname{spec} k_v$ for every place $v$. Then for almost all*[4] *$v$, $G_v$ is unramified.*

I thank Brian Conrad for pointing out the following proof to me.

*Proof (sketch).* We can fix a model for $G$ over the ring of $S$-integers $\mathfrak{o}_S$, where $S$ is some finite number of places. (In other words, if we write equations for $G$ over $k$, we can make sure that $S$ contains all the primes dividing the denominators, so that the coefficients are $S$-integers.) This can be done in a way that all fibers are connected.

[2] Since the map $H^1(\mathrm{Inn}(G)) \to H^1(\mathrm{Aut}(G))$ is not always injective, there is some ambiguity as to what "inner form" really means. Colloquially, it is often used to refer to a form, or an equivalence class of forms, of $G$ which arises by an inner twist (and hence parametrized by the image of this map). However, a more proper notion is one that would contain information about the "inner twist", so that isomorphism classes of inner forms are parametrized by $H^1(\mathrm{Inn}(G)$, and not by its image in $H^1(\mathrm{Aut}(G))$. "Inner twists" have a geometric explanation: the set $H^1(\mathrm{Inn}(G))$ parametrized isomorphism classes of $G_{\mathrm{ad}}$-*torsors* (i.e. principal homogeneous spaces), where $G_{\mathrm{ad}}$ is the adjoint group of $G$; for such a $G$-torsor $T$, the associated form of $G$ is the group $\mathrm{Aut}_G(T)$.

[3] A global field is a number field in characteristic zero, and the function field of a curve over a finite field in positive characteristic. Its *places* are the equivalence classes of norms on the field. In the function field case, these coincide with the (scheme-theoretic) closed points of the curve – the curve assumed non-singular, and are all *non-archimedean*. In the number field case the places include *non-archimedean* and a finite number of *archimedean* ones.

[4] "For almost all places" means "for all but a finite number".

For a finite set $S' \supset S$, all fibers away from $S'$ are reductive and split over an unramified extension. Indeed, to show that almost all fibers are reductive we may do this after a finite separable extension of $E/k$, and since every reductive group splits over a finite separable extension we may choose $E$ so that $G$ is split over $E$. Notice that $E$ is unramified almost everywhere over $k$, so the claim that $G$ splits over an unramified extension will be automatic. There is an explicit split reductive model of $G$ over $\mathbb{Z}$ (the "Chevalley model"), and since any two models agree at almost all places, this means that almost all fibers of the given model were reductive.

Now we will appeal to the existence and smoothness of "the scheme of Borel subgroups" of $G$ in order to show that at all places outside of $S'$ the group is quasisplit (hence unramified). Namely, given a reductive group $G$ over $\mathfrak{o}_{S'}$, there is an $\mathfrak{o}_{S'}$-scheme $\mathcal{B}$, the "scheme of Borel subgroups", representing the functor which to every $\mathfrak{o}_{S'}$-scheme $T$ associates the set of Borel subgroups of $G \times_{\mathrm{spec}\, \mathfrak{o}_{S'}} T$. The existence of this scheme is proven in [EGA 3.XXVI.3], together with the fact that it is projective, with geometrically integral fibers and *smooth* over $\mathfrak{o}_{S'}$. (Of course, if $G$ is reductive and quasi-split over a field $k$ then the scheme is isomorphic to $B\backslash G$, for some Borel $B$, since all Borel subgroups are conjugate and self-normalizing.)

Smoothness implies that for a dvr $R$, any point (Borel subgroup) over the residue field lifts to an $R$-point (Borel subgroup). Apply this to $R =$ the localization of $\mathfrak{o}_{S'}$ at any (nonzero) prime. Over the residue field $\mathbb{F}$, any (connected) reductive group is quasi-split; this follows from our discussion of forms: every form is an inner form of a quasi-split group $G'$, but $H^1(\mathrm{Inn}(G'))$ is trivial over a finite field because of *Lang's theorem* which states that $H^1$ of connected algebraic groups over finite fields is trivial. Thus, $G$ has a Borel over $\mathbb{F}$, and hence over $R$.

□

## 22.5 The dual group

A basic principle of the Langlands conjectures is that representations (and automorphic representations) of groups which are isomorphic to each other over the algebraic closure should be related. This is expressed by a conjectural parametrization of the representations by means of a *dual group*, which is a combinatorial construction based on the root datum of the group and the Galois action on it. (The next logical step is to deduce that if there is a map between dual groups then there should be a map between representations – of groups which a priori have nothing to do with each other. This is the principle of *functoriality*, but it is too early to get into this.) Non-combinatorial constructions of the dual group exist only in the *Geometric Langlands Program*.

To each reductive $k$-group $G$ we will associate a group of the form ${}^L G = \check{G} \rtimes \Gamma$, where $\check{G}$ is a complex reductive group and $\Gamma = \mathrm{Gal}(\bar{k}/k)$. The group ${}^L G$ is called the *Langlands dual group* or *L-group* of $G$.

The definition of $\check{G}$ is two phrases: One takes the root datum of $G$ and

inverts it. In other words, if $\Psi = (X, \Phi, \check{X}, \check{\Phi})$ is the root datum of $G$, that of $\check{G}$ will be $\check{\Psi} = (\check{X}, \check{\Phi}, X, \Phi)$. As we have seen, this uniquely determines a reductive group over an algebraically closed field. This group, without the Galois action, will sometimes be called the "dual" group of $G$, although when it is clear from the context we may just say "dual" for "Langlands dual". Notice that $\check{G}$ comes with a canonical maximal torus $A^*$.

Notice that the Dynkin diagram of the dual group is that of the group, with the arrows reversed, i.e. short roots become long and long roots become short. This, of course, does not fully determine the group. From the root datum we see that if $G$ is semisimple and simply connected then $\check{G}$ is adjoint and vice versa.

Examples (the isomorphism class of $G$ is given over the algebraic closure; since we are just applying an involution to the root data, we don't need to specify which one is $G$ and which one is $\check{G}$): $\mathrm{GL}_n \leftrightarrow \mathrm{GL}_n$, $\mathrm{SL}_n \leftrightarrow \mathrm{PGL}_n$, $\mathrm{SO}_{2n+1} \leftrightarrow \mathrm{Sp}_{2n}$, $\mathrm{SO}_{2n} \leftrightarrow \mathrm{SO}_{2n}$, $G_2 \leftrightarrow G_2$ etc.

To define the homomorphism $\Gamma \to \mathrm{Aut}(\check{G})$ defining the semi-direct product it is better to start from the split form $G^{\mathrm{spl}}$ of $G$. The Galois action on the based root datum $\Psi^+$ of $G^{\mathrm{spl}}$ is trivial, therefore $H^1(\mathrm{Aut}(\Psi^+)) = (\mathrm{Hom}(\Gamma, \mathrm{Aut}(\Psi^+))$ mod conjugation). Let now $\phi \in \mathrm{Hom}(\Gamma, \mathrm{Aut}(\Psi^+))$ a representative for the conjugacy class of the element corresponding to $G$ (more precisely: to the inner class of $G$). This defines the semidirect product: ${}^L G = \check{G} \rtimes \Gamma$ uniquely up to inner automorphisms.Notice that all forms in an inner class have the same $L$-group. We will see (much later) that the best way to formulate the Langlands conjectures deals with all inner forms at the same time, not each one separately.

*Remark.* As a matter of convenience, since $\Gamma$ acts on $\mathrm{Aut}(\check{G})$ through a quotient $\mathrm{Gal}(F/k)$, we often use $\check{G} \rtimes \mathrm{Gal}(F/k)$ instead of ${}^L G$. We will adopt this habit, and feel free to refer to $\check{G} \rtimes \mathrm{Gal}(F/k)$ (which, of course, is not uniquely defined) as the Langlands dual group whenever this causes no confusion. For instance, if $G$ is split we will usually use $\check{G}$.

## 22.6 Basic examples

### 22.6.1

Let $F$ be a Galois extension of $k$ and $G$ a split group over $F$. Let $G' = \mathrm{Res}_{F/k} G$. Then ${}^L G' = (\check{G})^{F \hookrightarrow \bar{k}} \rtimes \mathrm{Gal}(F/k)$ (s. the previous remark). (Exercise!)

### 22.6.2 Inner forms of $\mathrm{GL}_n$.

Consider $\mathrm{GL}_n$ as a subvariety of $\mathrm{Mat}_n$. Every inner automorphism of $\mathrm{GL}_n$ extends to $\mathrm{Mat}_n$, therefore defines a form of the latter, a *central simple algebra* $D$ of dimension $n^2$ over $k$. The corresponding inner form of $\mathrm{GL}_n$ is $G = D^\times$ (the multiplicative group of invertible elements of $D$, regarded as a variety over $k$). Vice versa, any automorphism of a simple algebra is inner (a special case of

the *Skolem-Noether theorem*), hence all central simple algebras correspond to inner forms of $\mathrm{GL}_n$.

Notice the following part of the long exact sequence:

$$H^1(\mathrm{GL}_n) \to H^1(\mathrm{PGL}_n) \to H^2(\mathbb{G}_m)$$

corresponding to the short exact sequence:

$$1 \to \mathbb{G}_m \to \mathrm{GL}_n \to \mathrm{PGL}_n\,.$$

Since $H^1(\mathrm{GL}_n) = 1$ (Hilbert's theorem 90), $H^1(\mathrm{PGL}_n = \mathrm{Inn}(\mathrm{GL}_n))$ injects into $\mathrm{Br}_k = H^2(\mathbb{G}_m)$, the Brauer group of $k$. Hence for every central simple element of dimension $n^2$ we get an element of the Brauer group, and this way we could recover the classification of central simple algebras up to equivalence.

Central simple algebras of dimension 4 are called *quaternion algebras*. In characteristic other than two, all isomorphism classes admit a (non-unique) presentation of the form: $D = Q(a,b) := k\{i,j\}/\langle i^2 = a, j^2 = b, ij = -ji\rangle$. (Angular brackets denote the free non-commutative algebra.) Caution: This includes *all* central simple algebras of dimension 4, i.e. both the division algebras *and* $\mathrm{Mat}_n$. Whether a quaternion algebra is *split* (=isomorphic to $\mathrm{Mat}_n$) or not depends on $a$ and $b$. We will discuss this in more detail later, when we recall class field theory and the Hilbert symbol.

The structure of quaternion algebras is very important and very educational for number theorists. An excellent reference is: M.F. Vigneras, *Arithmetique des algèbres de quaternions*.

### 22.6.3 Unitary groups.

*Exercise.* For every $n$, $\mathrm{Out}(\mathrm{GL}_n) = \mathbb{Z}/2$, represented by $g \mapsto {}^t g^{-1}$.

Let $G = \mathrm{Res}_{E/F}\mathrm{GL}_n$, where $E/F$ is a quadratic (Galois) extension. Then $G(\bar F) = \mathrm{GL}_n(\bar F)\times\mathrm{GL}_n(\bar F)$ and the Galois action is given as follows: If ${}^\gamma(g_1,g_2) = \tau(\gamma)({}^\gamma g_1, {}^\gamma g_2)$ where $\tau(\gamma)$ is the trivial or the non-trivial permutation of the two factors according as $\gamma \equiv 1$ in $\mathrm{Gal}(E/F)$ or not. Notice that $G$ comes with an $F$-linear involution, from the action of $\sigma$ on $\mathrm{GL}_n(E)$, which in order to distinguish from the Galois action on points of $G$ we will denote by $\overline{(\ )}$: $\overline{(g_1,g_2)} = (g_2,g_1)$.

Hence $\check G = \mathrm{GL}_n \times \mathrm{GL}_n$. The homomorphism $\Gamma = \mathrm{Gal}(\bar F/F) \to \mathrm{Aut}(\check G)$ (unique up to $\mathrm{Inn}(A^*)$) which defines the $L$-group factors through $\mathrm{Gal}(E/F)$ and the non-trivial element $\sigma$ of $\mathrm{Gal}(E/F)$ maps to the automorphism $(g_1,g_2) \mapsto (g_2,g_1)$. This defines ${}^L G$, which for simplicity we present as $(\mathrm{GL}_n(\mathbb{C})\times\mathrm{GL}_n(\mathbb{C}))\rtimes\{1,\sigma\}$.

Let $H$ be a non-degenerate hermitian matrix, and let $U = U_n \subset G$ be the unitary group of this hermitian form, i.e. the group of $g \in G$ such that ${}^t\bar g H g = H$. Clearly, $U_E \simeq \mathrm{GL}_n$.

**Claim.** *All unitary groups in $n$ variables are inner forms of each other.*

Indeed, all non-degenerate hermitian forms are $G(\bar F)$-conjugate, so if $U'$ is another unitary group in $G$ then $U'(\bar F) = gU(\bar F)g^{-1}$ for some $g \in G(\bar F)$, and

the cocycle $c : \Gamma \to \mathrm{Aut}(U)$ corresponding to $U'$ is given by $c(\gamma) = \mathrm{Inn}(g \cdot {}^{\gamma}g^{-1})$, and $g \cdot {}^{\gamma}g^{-1} \in U(\bar{F})$. Notice that $\gamma \mapsto g \cdot {}^{\gamma}g^{-1}$ is a coboundary in $G$, but not (necessarily) in $U$.

(Vice versa, one can show that all inner forms of $U$ are "generalized unitary groups", namely they arise as follows: There is a central division algebra $D$ over $E$ with an $F$-linear involution "of the second kind", i.e. the involution restricted to $E$ (the center of $D$) is non-trivial, to be denoted again by $\overline{(\ )}$. Now we define a "generalized unitary group" in the same way as above, but using $\mathrm{Mat}_m(D)$ instead of $\mathrm{Mat}_n(E)$, where $m^2 \cdot \dim(D/E) = n^2$.)

We now compute the dual group of $U$. We have $\check{U} = \mathrm{GL}_n$. To determine the action of the Galois group, choose the hermitian matrix $H = I_n$ and the isomorphism: $(\mathrm{GL}_n)_{\bar{F}} \to U_{\bar{F}}$ given by $g \mapsto (g, {}^t g^{-1}) \in U \subset G$. Then one sees that the homomorphism $\Gamma \to \mathrm{Aut}(\check{U})$ factors through $\mathrm{Gal}(E/F)$ and $\sigma \mapsto (g \mapsto {}^t g^{-1})$. Hence ${}^L U(\mathbb{C}) = \mathrm{GL}_n(\mathbb{C}) \rtimes \{1, \sigma\}$ with this action.

# Chapter 23

# Brauer groups, Galois cohomology.

## 23.1 References.

- J. Milne, *Class field theory.*
- J.-P. Serre, *Galois cohomology.*

## 23.2 Central simple algebras.

Let $k$ be a field, an algebra $A$ over $k$ is *simple* if it has no two-sided ideals other than 0 and $A$, and central if its center is $A$. By Wedderburn's theorem, every finite-dimensional simple algebra is isomorphic to $\mathrm{Mat}_n(D)$, where $D$ is a division algebra over $k$, and this is central iff $D$ is central. We also recall the Skolem-Noether theorem: If $A, B$ are simple, finite dimensional algebras over $k$ and $B$ is central then any non-zero (hence automatically injective) homomorphisms: $A \to B$ are conjugate by a unit of $B$ (i.e. they coincide after applying an inner automorphism to $B$).

From now on, all our algebras will be finite-dimensional over $k$. The tensor product of two central simple algebras is again central simple. We define an equivalence relation $A \sim B$ on central simple algebras as follows: $A \sim B \iff A \otimes \mathrm{Mat}_m \simeq B \otimes \mathrm{Mat}\, n$ for some $m, n$. We define $\mathrm{Br}(k)$ to be the set of isomorphism classes of central simple algebras modulo this equivalence relation, and denote by $[A]$ the class of $A$. Tensor product descends to provide $\mathrm{Br}(k)$ with the structure of a monoid. In fact it is a group: This is proven by the relation $A \otimes A^{\mathrm{opp}} \simeq \mathrm{End}_k(V)$, where $A^{\mathrm{opp}}$ the 'opposite' algebra of $A$. (Proof: $A \otimes A^{\mathrm{opp}}$ acts $k$-linearly on $A$ by 'left and right multiplication'; by simplicity, it injects into $\mathrm{End}_k(A)$, and by dimension counting, it surjects.) Hence the name *Brauer group.*

Let $E/k$ be a finite Galois extension, and $A$ a central simple algebra over $k$. We say that $A$ *splits over* $E$ if $[A \otimes E] = 1$ in $\mathrm{Br}(E)$ (i.e. it is isomorphic to a matrix algebra over $E$). We define $\mathrm{Br}(E/k)$ to be the subgroup of $\mathrm{Br}(k)$ consisting of elements which split over $E$. (One can easily see that the previous sentence makes sense.)

**Facts.**

- *There is a natural isomorphism of abelian groups:*

$$\mathrm{Br}(E/k) \simeq H^2(\mathrm{Gal}(E/k), E^\times).$$

- *If $L/E/k$ is a tower of Galois extensions, then the injection:* $\mathrm{Br}(E/k) \hookrightarrow \mathrm{Br}(L/k)$ *corresponds to the so-called 'inflation-restriction' exact sequence in group cohomology:*[1]

$$0 \to H^2(\mathrm{Gal}(E/k), E^\times) \to H^2(\mathrm{Gal}(L/k), L^\times) \to H^2(\mathrm{Gal}(L/E), L^\times).$$

- *Given these injections,* $\mathrm{Br}(k) = H^2(\mathrm{Gal}(\bar{k}/k), \mathbb{G}_m)$.

The element of $H^2$ corresponding to the class of a central simple algebra of dimension $n^2$ can be obtained[2] by the boundary map:

$$H^1(\mathrm{PGL}_n) \hookrightarrow H^2(\mathbb{G}_m)$$

of the long exact cohomology sequence corresponding to the exact sequence:

$$1 \to \mathbb{G}_m \to \mathrm{GL}_n \to \mathrm{PGL}_n \to 1.$$

Recall from the discussion of inner forms of $\mathrm{GL}_n$ that $H^1(\mathrm{PGL}_n)$ parametrizes isomorphism classes of central simple algebras of dimension $n^2$. You will understand this long exact sequence better after you read the rest of this section.

## 23.3 Abelian and non-abelian Galois cohomology

If $\Gamma$ is a topological group acting on a group $A$ (considered discrete), we saw in lecture 2 an explicit definition of the first cohomology $H^1$, which we recall here: First we define the set of 1-cocycles as the set of (continuous) maps: $c : \Gamma \to A$ satisfying the condition:

$$c(\gamma_1\gamma_2) = c(\gamma_1) \cdot {}^{\gamma_1}c(\gamma_2).$$

The group $A$ acts on cocycles by ${}^b c(\gamma) = bc'(\gamma)^\gamma b^{-1}$, and we define $H^1(\Gamma, A)$ to be the set of orbits for this action. We discussed that this is a pointed set.

[1] An assumption for the 'inflation-restriction' sequence to be exact on the left is that the corresponding $H^1$ groups are trivial, which is Hilbert's theorem 90.

[2] I think that the isomorphism $\mathrm{Br}(k) \simeq H^2(\mathbb{G}_m)$ is defined to be the inverse of this in some parts of the literature.

If $A$ is an abelian group, then $H^1$ has the natural structure of an abelian group, moreover there is a much better definition which immediately defines groups $H^i$ for every positive $i$: $H^i(\Gamma, A)$ is $i$-th derived functor of the left-exact functor $H^0(\Gamma, A) = A^\Gamma$ ($\Gamma$-invariants). (Moreover, one can define the homology groups $H_i(\Gamma, A)$ as the $i$-th derived functor of the right-exact functor: $H_0(\Gamma, A) = A_\Gamma$ (co-invariants). We do not discuss here the notion of derived functors, suffice it to say that they are canonically defined and that given a short exact sequence $0 \to A \to B \to C \to 0$ of $\Gamma$-modules we get corresponding long exact sequences of cohomology and homology groups. For non-abelian groups there is no such 'derived functor' construction, and even defining a reasonable analog for $H^2$ in cases where we need it is very tricky.

The point now is that these cohomology groups are abstractly defined in terms of the group and the module, but in practice they have many reincarnations which make them useful and help them compute them. As an example, we mentioned above how $H^2(\mathrm{Gal}(E/k), E^\times)$ is interpreted as a Brauer group; using this incarnation one can prove certain facts about this group; and then use this fact to understand something about a different incarnation of this group. This happens, for instance, in a certain approach to local class field theory. The goal of the discussion below is to collect several basic facts about Galois cohomology groups of algebraic groups and discuss some basic incarnations.

As a matter of notation, if $\Gamma$ is the absolute Galois group of some field $k$ and $G$ is an algebraic group over $k$, we use the notation $H^i(\Gamma, G)$ to denote the group $H^i(\Gamma, G(\bar{k}))$. (We write $\bar{k}$ but we mean the separable closure. Recall the "continuous" assumption: we consider $G(\bar{k})$ as a discrete group, so every cocycle factors through a finite quotient of the Galois group.) Sometimes we also write just $H^i(G)$.

## 23.4 Basic and important facts of Galois cohomology

In the formulations below, notice that even though we are stating things for infinite Galois groups, since all cohomology groups are defined using *continuous* cocycles and the modules are considered *discrete* they are really limits of the corresponding cohomology groups over all finite extensions of the field.

*The additive group:* The additive group $\mathbb{G}_a$ is *cohomologically trivial*: $H^i(\mathbb{G}_a) = 0$ for every $i > 0$. This follows from the following facts:

**Lemma 4.1.** *Let $\Gamma$ be a finite group, $V$ an abelian group and $\tilde{V} = \mathrm{Ind}_{\{1\}}^\Gamma V$. Then $\tilde{V}$ is cohomologically trivial as a $\Gamma$-module.*

**Theorem 4.2** (Normal Basis Theorem)**.** *If $E/k$ is a finite Galois extension with Galois group $\Gamma$ then there exists $a \in E$ such that $\Gamma \cdot a$ is a basis for $E$ over $k$. Hence, as a $k[\Gamma]$-module, $E \simeq k[\Gamma] = \mathrm{Ind}_{\{1\}}^\Gamma k$.*

*The multiplicative group:* For any field $k$ we have $H^1(\mathbb{G}_m) = 1$. This is *Hilbert's theorem 90*. However, higher cohomology groups are, in general, non-zero; for example, $H^2(\mathbb{G}_m)$ is the *Brauer group* of $k$.

$\mathrm{GL}_n$ *and* $\mathrm{SL}_n$*:* A generalization of the previous statement is that $H^1(\mathrm{GL}_n) = 1$. This is also referred to as Hilbert's theorem 90. Beware: this *is not* a statement about forms of $\mathrm{GL}_n$ because if we change the Galois action we also change the cohomology groups. By the short exact sequence: $1 \to \mathrm{SL}_n \to \mathrm{GL}_n \to \mathbb{G}_m \to 1$ and the surjectivity of the 'determinant' map we also get that $H^1(\mathrm{SL}_n) = 1$.

*Cohomological dimension:* A group $\Gamma$, or a field $k$ with absolute Galois group is said to have *cohomological dimension* $\leqslant n$ if for every *torsion* $k$-module $A$ we have $H^i(\Gamma, A) = 0$ for $i > n$.

*Finite fields:* Fact: $\mathrm{cd}(\widehat{\mathbb{Z}}) = 1$. Corollary: for a finite field $k$, $H^i(k, \mathbb{G}_m) = 1$ for every $i$. For instance, there are no non-trivial central simple algebras over $k$.

*Lang's theorem:* Let $k$ be a field of cohomological dimension $\leqslant 1$ (such as a finite field) and let $G$ be a *connected* algebraic group over $k$, then $H^1(k, G) = 1$.

*Torsors:* A *torsor* or *principal homogeneous space* for an algebraic group $G$ over $k$ is an algebraic variety $X$ with a $G$-action such that the action of $G(\bar{k})$ on $X(\bar{k})$ is simply transitive. (Hence, $X(\bar{k}) = G(\bar{k})$ as a $G(\bar{k})$-space non-canonically, since there is no 'preferred' point on $X$.) Clearly, the torsor is trivial (i.e. $k$-isomorphic as a $G$-space to $G$) if and only if $X(\bar{k}) \neq \varnothing$.

**Fact.** *(easy to check! exercise!): Isomorphism classes of $G$-torsors over $k$ are classified by $H^1(k, G)$. This is a canonical isomorphism since there is a preferred class trivial torsors, which will correspond to the distinguished point of $H^1(k, G)$.*

*Homogeneous spaces:* Let $G$ be an algebraic group over $k$. A *homogeneous space* for $G$ over $k$ is a variety $X$ over $k$ with a (let's say right) $G$ action such that $G(\bar{k})$ acts transitively on $X(\bar{k})$. Assume that $X(k) \neq \varnothing$, and pick a $k$-point $x_0$. Then its stabilizer will be a closed algebraic subgroup $H$ defined over $k$, and the orbit map $g \mapsto x_0 \cdot g$ defines an isomorphism of varieties: $X \simeq H\backslash G$.

**Fact.** *(check it as well!): The set of $G(k)$-orbit on $X(k)$ is parametrized by the kernel of:*

$$H^1(k, H) \to H^1(k, G).$$

*This parametrization depends on the choice $x_0$ – its $G(k)$ orbit will correspond to the distinguished point.*

*Remark.* The above fact can be seen as a "long exact sequence":

$$1 \to H^0(H) \to H^0(G) \to H^0(X) \to H^1(H) \to H^1(G).$$

*Application to vector spaces with tensors:* Let $V$ be a $k$-vector space of dimension $n$ and $Q$ be a tensor of type $(p,q)$ for $V$, i.e. an element of $V^p \otimes (V^*)^q$. Naturally, the group $\mathrm{GL}_V$ acts on the space of such tensors (let's say on the right), and we would like to classify over $k$ the tensors $Q'$ which are $\bar{k}$-isomorphic to $Q$, i.e. such that $Q' \in Q \cdot \mathrm{GL}_V(\bar{k})$.

Examples: A quadratic form is a symmetric $(0,2)$-tensor and non-degenerate quadratic forms are a homogeneous space under $\mathrm{GL}_V$; the stabilizers are the *orthogonal groups* $O_n$. Similarly a *symplectic form* is a non-degenerate (here $n = \dim V$ must be even) alternating $(0,2)$-tensor, and all such form a homogeneous space for $\mathrm{GL}_V$ with stabilizers the *symplectic groups* $\mathrm{Sp}_n$.

Let $G_Q \subset \mathrm{GL}_V$ be the stabilizer of $Q$, then $k$-equivalence classes (i.e. $GL_V(k)$-orbits) of such $Q'$ are parametrized by:

$$\ker(H^1(k,Q) \to H^1(k,\mathrm{GL}_V)) = H^1(k,Q)$$

by Hilbert's theorem 90. For instance we have parametrizations (depending on base points):

$$\left\{ \begin{array}{c} k-\text{equivalence classes of} \\ \text{non-degenerate quadratic forms} \\ \text{in } n \text{ variables} \end{array} \right\} \leftrightarrow \quad H^1(k, O_n).$$

and:

$$\left\{ \begin{array}{c} k-\text{equivalence classes of} \\ \text{symplectic forms} \\ \text{in } n \text{ variables} \end{array} \right\} \leftrightarrow \quad H^1(k, \mathrm{Sp}_n).$$

It is known that any two symplectic forms are $k$-isomorphic, hence:

$$H^1(k, \mathrm{Sp}_n) = 1.$$

*Cohomology groups over a p-adic field:* The following generalizes Tate-Nakayama duality for tori, gives a very convenient way of computing cohomologies over a $p$-adic field, and the (somewhat surprising, at first glance) fact that in this case $H^1$ is actually a group![3]

**Theorem 4.3** (Kottwitz)**.** *Let $k$ be a local non-archimedean field, $G$ a connected reductive group over $k$ and ${}^L G = \check{G} \rtimes \Gamma$ its Langlands dual (notation as previously). Then there is a canonical isomorphism:*

$$H^1(k,G) = \left(\pi_0(\mathcal{Z}(\check{G})^\Gamma)\right)^*$$

*where by $\mathcal{Z}$ we denote the center, by $\pi_0$ the group of connected components and by an asterisk the dual (i.e. character) group.*

As an exercise, compute cohomologies for several groups, including tori, and verify that the computation coincides with things you know, such as the index $(k^\times : N_k^E E^\times)$ where $E/k$ is a finite Galois extension.

[3] The group structure arises from the group structure on $H^1(k,T)$, where $T$ is any maximal anisotropic torus in $G$.

## 23.5 Reciprocity for global Brauer groups.

Recall (from the next chapter!) that by local class field theory we have canonical isomorphisms:

$$\mathrm{Br}(k) \simeq \mathbb{Q}/\mathbb{Z}$$

for any local non-archimedean field, and:

$$\mathrm{Br}(\mathbb{R}) \simeq \frac{1}{2}\mathbb{Z}/\mathbb{Z}.$$

Our purpose here is to explain the following theorem and its implications. The theorem is closely connected with the proof of the reciprocity law of global class field theory, but we won't describe the connection.

**Theorem 5.1.** *Let $k$ be a global field. There is a canonical short exact sequence:*

$$0 \to \mathrm{Br}(k) \to \oplus_v \mathrm{Br}(k_v) \to \mathbb{Q}/\mathbb{Z} \to 0 \tag{23.1}$$

*(called the* fundamental exact sequence *of global class field theory. The product here ranges over all completions of $k$ (including archimedean ones).*

It is easy to understand the first map componentwise (as the restriction map to decomposition groups). The fact that its image lies in the *direct sum* is non-trivial, but not hard. It is similar to our discussion of why every reductive group is quasi-split almost everywhere. In this case it implies:

> A central simple algebra over $k$ splits at almost every place.

Injectivity of the first map is the *Hasse principle for Brauer groups*. The name "Hasse principle" for such a property is understood through the following (but see also the paragraph on the Hasse principle below):

**Corollary 5.2.** *Let $A, B$ be two central simple algebras of the same dimension over a global field $k$. If $A_v$ (the completion of $A$ at $v$, i.e. $A \otimes_k k_v$) is isomorphic to $B_v$ at every $v$ then $A$ and $B$ are isomorphic over $k$.*

The map on the right comes from adding the "invariant" maps (i.e. the aforementioned isomorphisms). Its surjectivity implies that we can describe a central simple algebra over the global field $k$ by just describing its local components! For instance, the local "invariants" of quaternion algebras are either $1/2$ or zero. Hence:

> For every finite, even collection of places $S$ which do not contain complex places, there is *precisely one* quaternion algebra over $k$ which is ramified exactly at the places in $S$.

## 23.6 The Hasse principle.

We say that a group $G$ satisfies the *Hasse principle* if the canonical map of Galois cohomologies:

$$H^1(k, G) \to \prod_v H^1(k_v, G_v)$$

is injective. Notice that by Kottwitz's isomorphism for $H^1(k_v, G_v)$ if $v$ is finite, if $G$ is semisimple simply connected then the product on the right is only over the infinite places. We have:

**Theorem 6.1** (Kneser, Harder, ...)**.** *Let $G$ be a semisimple, simply connected group, then it satisfies the Hasse principle.*

Let us see what such a statement has to do with the classical Hasse principle for quadratic forms. Recall that the theorem of Hasse-Minkowski states that:

> A quadratic form in $n$ variables has a non-trivial zero over $k$ if it has a non-trivial zero over $k_v$, for every $v$.

By an easy argument involving Witt't theorem for quadratic forms, this is equivalent to the following:

> Two quadratic forms $Q_1$ and $Q_2$ are isomorphic over $k$ if and only if they are isomorphic over $k_v$, for every $v$.

As we have seen, isomorphism classes of non-degenerate quadratic forms are classified (depending on choice of a 'base point') by elements in $H^1(k, O_n)$. Therefore, the Hasse-Minkowski theorem (we can assume non-degeneracy) is equivalent to:

**Theorem 6.2** (Hasse-Minkowski)**.** *The canonical map $H^1(k, O_n) \to \prod_v H^1(k_v, O_{n,v})$ is injective.*

(Notice that $O_n$ is not simply connected, so this is not a special case of the previous one.)

The question of whether a Hasse principle holds or not is a very important and often difficult one. If you want to enrich your vocabulary, look at a book on elliptic curves to learn the definition of the *Tate-Shafarevich group*: it is an ostruction to the Hasse principle.

# Chapter 24

# Recollection of class field theory.

Class field theory, the cornerstone of algebraic number theory developed in the first half of the 20th century, is a subject that one should study by itself and for its own sake, and it preceeds the theory of automorphic representations, both logically and as a necessary tool. Therefore, the notes for this lecture are not intended to present the subject in any complete fashion, but only to gather the basic results that will be needed in our course.[1]

## 24.1 References.

- Cassels and Fröhlich (eds.), *Algebraic number theory.*
- S. Lang, *Algebraic number theory.*
- J. Milne, *Class field theory*, on the web: `http://www.jmilne.org/`.
- J.P. Serre, *Local fields.*

Here we mostly follow Milne's notes.

## 24.2 Local class field theory.

Let $k$ be a *local field*. The term *local field* will be reserved throughout the notes for a locally compact one. Hence, a local field is either $\mathbb{R}$ or $\mathbb{C}$ –called *archimedean*– or the quotient field of a discrete valuation ring with finite residue field –called *non-archimedean*. The main results of local class field theory are:

[1] Here it is appropriate to recall a quote of Prof. James Milne during a lecture of his at the 2003 Summer School on Automorphic Forms in Toronto: "I have fifteen minutes left and two things to talk about; one of them is class field theory."

**Theorem 2.1** (Local Reciprocity Law)**.** *There is a unique homomorphism (*local Artin map *or* local reciprocity map*)*

$$\phi_k : k^\times \to \mathrm{Gal}(\bar{k}^{\mathrm{ab}}/k)$$

*with the following properties:*

1. *If $k$ is non-archimedean, for any prime element $\varpi$ of $k$ and any finite unramified extension $E$ of $k$,*

$$\phi_k(\varpi)|_E = \mathrm{Frob}_{E/k}\,.$$

2. *For any finite extension $E$ of $k$, the composition $k^\times \to \mathrm{Gal}(\bar{k}/k)^{\mathrm{ab}} \to \mathrm{Gal}(E/k)^{\mathrm{ab}}$ induces an isomorphism:*

$$\phi_{E/k} : k^\times/N_k^E E^\times \simeq \mathrm{Gal}(E/k)^{\mathrm{ab}}.$$

As formulated, the theorem contains the *Norm Limitation Theorem*, which states that if $E/k$ is Galois and $F \subset E$ is the maximal abelian subextension then $N_k^E E^\times = N_k^F F^\times$.

**Theorem 2.2** (Local Existence Theorem)**.** *A subgroup $N$ of $k^\times$ is of the form $N_k^E E^\times$ for some finite abelian extension $E^\times$ of $k$ if and only if it is of finite index and open.*

Putting together the reciprocity and the existence theorem, we get an isomorphism:[2]

$$\mathrm{Gal}(\bar{k}/k)^{\mathrm{ab}} \simeq \widehat{k^\times} \tag{24.1}$$

where $\widehat{k^\times}$ denotes the *profinite completion* of $k^\times$.

We reformulate the local reciprocity Law in terms of Galois cohomology: First, we recall the *Tate cohomology groups* – a construction which joins together the long exact sequence for group cohomology – the derived functor for $A \mapsto A^G$ – and homology – the derived functor for $A \mapsto A_G$. We assume that the group $G$ is finite, then for any $G$-module $A$ the group $H_T^i(A)$ is defined as:

$$H_T^i(G,A) = \begin{cases} H^i(G,A), & \text{if } i \geqslant 1, \\ A^G/N_G(A), & \text{if } i = 0, \\ \ker(N_G)/I_G(A), & \text{if } i = -1 \text{ and} \\ H_{-i-1}(A), & \text{if } i < -1. \end{cases}$$

Here $N_G$ denotes the norm map $a \mapsto \sum_G ga$, and $I_G$ the *augmentation ideal* of $\mathbb{Z}[G]$ generated by the elements of the form $(1-g), g \in G$. (Hence the coinvariants $A_G = A/I_G(A)$.)

[2] We formulate the statements for the case of characteristic zero, but they also hold for function fields if we replace the algebraic closure by the separable closure.

The Tate groups have the benefit that they join the two long exact sequences, i.e. a short exact sequence of modules:

$$0 \to A \to B \to C \to 0$$

gives rise to a two-way long exact sequence of Tate groups:

$$\dots H_T^i(A) \to H_T^i(B) \to H_T^i(C) \to H_T^{i+1}(A) \to \dots.$$

Now we see what the isomorphism of class field theory says in cohomological language. Let $E/k$ be a finite Galois extension and let $G = \mathrm{Gal}(E/k)$, then:

$$\phi_{E/k} : \mathrm{Gal}(E/k)^{\mathrm{ab}} = H_T^{-2}(G, \mathbb{Z}) \simeq H_T^0(G, E^\times) = k^\times / N_k^E E^\times.$$

This isomorphism will be provided by *cup product* by a canonical element $u_{E/k}$ of $H^2(G, E^\times)$. Hence, the proof of the local existence theorem rests upon proving the following two facts (we assume, of course, that $k$ is non-archimedean):

1. The first fact is of arithmetic nature: $H^2(\mathrm{Gal}(\bar{k}/k), \mathbb{G}_m) \simeq \mathbb{Q}/\mathbb{Z}$ canonically. Moreover, for any abelian extension $E$ of degree $n$ the image of $H^2(\mathrm{Gal}(E/k), E^\times)$ is the subgroup generated by $1/n$. The element of $H^2(\mathrm{Gal}(E/k), E^\times)$ corresponding to $1/n$ is called the *fundamental class* $u_{E/k}$ of the extension $E/k$.

2. The second fact is purely a statement in group cohomology and is called *Tate's theorem*: Let $G$ be a finite group and $A$ a $G$-module. Suppose that for all subgroups $H$ of $G$ we have $H^1(H, A) = 0$ and $H^2(H, A)$ is cyclic of order equal to $|H|$. Then, cup product by a generator $u \in H^2(G, A)$ defines for all $i$ an isomorphism:

$$H_T^i(G, \mathbb{Z}) \simeq H_T^{i+2}(G, A).$$

We discuss the first fact, which is of wider number-theoretic interest since, as we have already discussed, the group $H^2(\mathrm{Gal}(\bar{k}/k), \mathbb{G}_m)$ is the *Brauer group* of $k$ and classifies isomorphism classes of central simple algebras over $k$. One shows first that $H^2(\mathrm{Gal}(k^{\mathrm{ur}}/k), k^{\mathrm{ur}}) = \mathbb{Q}/\mathbb{Z}$. This is relatively easy to do; one shows that the units in every unramified extension are cohomologically trivial; hence for every finite unramified $E/k$ we have

$$H^2(\mathrm{Gal}(E/k), E^\times) = H^2(\mathrm{Gal}(E/k), E^\times/\mathfrak{o}_E^\times = \mathbb{Z})$$

(we normalize the isomorphism with $\mathbb{Z}$ to map $\varpi \mapsto 1$), and then the latter is equal to $H^1(\mathrm{Gal}(E/k), \mathbb{Q}/\mathbb{Z}) = \mathrm{Hom}(\mathrm{Gal}(E/k), \mathbb{Q}/\mathbb{Z})$ (by the long exact sequence of $0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0$), which is cyclic with canonical generator $\mathrm{Frob}_{E/k} \mapsto 1/\deg(E/k)$. The final step is to show that $H^2(\mathrm{Gal}(k^{\mathrm{ur}}/k), k^{\mathrm{ur}}) = H^2(\mathrm{Gal}(\bar{k}/k), \mathbb{G}_m)$, or: $H^2(\mathrm{Gal}(\bar{k}/k^{\mathrm{ur}}), \mathbb{G}_m) = 0$. One way to do that is by using its interpretation as Brauer group and showing that *every central simple algebra over $k$ splits over an unramified extension*, but there are also other ways.

We have outlined a proof of the local reciprocity theorem. For the existence theorem there are several proofs, e.g. using Lubin-Tate formal group laws (s. Milne) or softer methods using generalities about *class formations* together with Kummer and Artin-Schreier theory (s. Serre).

## 24.3 Global class field theory.

Let $E/k$ be a finite abelian Galois extension of global fields. Let $\Gamma = \mathrm{Gal}(E/k), \Gamma_v = \mathrm{Gal}(E_w/k_v)$ for every place $v$ of $k$ and some place $w$ of $E$ over $v$. Notice that since $E/k$ is abelian the decomposition groups of all $w/v$ are the same, so $\Gamma_v$, as a subgroup of $\Gamma$, does not depend on $w$. From local class field theory, we have homomorphisms: $k_v^\times \to \Gamma_v$ for all $v$. Moreover, for almost every $v$ these homomorphisms are trivial on $\mathfrak{o}_v^\times$, the units of $k_v$. Therefore, their product defines a homomorphism (*global Artin map* or *global reciprocity map*):

$$\mathbb{A}_k^\times = \mathbb{G}_m(\mathbb{A}_k) \to \Gamma.$$

The *Global Reciprocity Law* of class field theory states:

**Theorem 3.1.** *The subgroup $k^\times \subset \mathbb{A}_k^\times$ lies in the kernel of the reciprocity map. More precisely:*

$$\mathbb{A}_k^\times/(k^\times N_k^E \mathbb{A}_E^\times) \simeq \Gamma$$

*under the reciprocity map.*

And the *Global Existence Theorem* states:

**Theorem 3.2.** *For a subgroup $N \subset \mathbb{A}_k^\times/k^\times$ to be the kernel of the reciprocity homomorphism for some finite abelian extension $E$, it is necessary and sufficient that it be open of finite index.*

Putting together the reciprocity and the existence theorem, we get:

$$\mathrm{Gal}(\bar{k}/k)^{\mathrm{ab}} \simeq \widehat{\mathbb{A}_k^\times/k^\times}. \tag{24.2}$$

We will not discuss the proofs of global class field theory, but we will discuss several applications.

## 24.4 Hilbert symbols.

Let $n$ be a positive integer and let $k$ be a local field which contains the $n$-th roots of unity $\mu_n$. Then we have two theories of cyclic degree-$n$ extensions of $k$; the first is Kummer theory and the other is class field theory. *Hilbert symbols* are a way to compare the two theories.

Namely, the power-$n$ Hilbert symbol $(\ ,\ ) : k^\times \times k^\times \to \mu_n$ is defined as follows: Let $a, b \in k^\times$. Consider the Kummer extension $k(a^{1/n})$ and the Galois element $\phi_k(b)$. Then we define:

$$\phi_k(b)(a^{1/n}) = (a,b)a^{1/n}. \tag{24.3}$$

There is also a cohomological description of the Hilbert symbol. Let $\Gamma = \mathrm{Gal}(\bar{k}/k)$. Kummer theory, at this level, is the obvious cohomological statement that $k^\times/(k^\times)^n \simeq H^1(\Gamma, \mu_n) = \mathrm{Hom}(\Gamma, \mu_n)$ (coming from the long exact sequence for $1 \to \mu_n \to \mathbb{G}_m \to \mathbb{G}_m \to 1$, where the third arrow is raising to the $n$-th power). The Hilbert symbol is the map:

$$k^\times \times k^\times \to k^\times/(k^\times)^n \times k^\times/(k^\times)^n \simeq H^1(\Gamma, \mu_n) \times H^1(\Gamma, \mu_n) \to$$

$$\to H^2(\Gamma, \mu_n \otimes \mu_n) \simeq \mu_n.$$

Here, the second-last arrow is cup product and the last arrow is given by cup product $H^2(\Gamma, \mu_n) \times H^0(\Gamma, \mu_n) \simeq H^2(\Gamma, \mu_n \otimes \mu_n)$ and the canonical isomorphisms: $H^2(\Gamma, \mu_n) = \mathbb{Z}/n\mathbb{Z}$, $H^0(\Gamma, \mu_n) = \mu_n$ and $\mathbb{Z}/n\mathbb{Z} \otimes \mu_n = \mu_n$.

Here is an application of the Hilbert symbol: Recall that if the characteristic of $k$ is not 2, every quaternion algebra over $k$ admits a presentation $D = Q(a,b) := k\{i,j\}/\langle i^2 = a, j^2 = b, ij = -ji\rangle$ for some $a, b \in k$. If $k$ is a local field the algebra is split if and only if the quadratic (i.e. $n = 2$) Hilbert symbol $(a,b)$ is trivial. (More precisely, the class $[D]$ in the two-torsion ($\simeq \mathbb{Z}/2$) of $\mathrm{Br}(k)$ is non-trivial if and only if $(a,b)$ is so.) Notice that the quadratic Hilbert symbol admits the following equivalent definition:

$$(a,b) = \begin{cases} 1, & \text{if } z^2 = ax^2 + by^2 \text{ has a non-zero solution,} \\ -1, & \text{otherwise.} \end{cases}$$

Now let $k$ be a global field containing the $n$-th roots of unity, $a, b \in k$. For every place $v$ of $k$ we have the power-$n$ Hilbert symbol $(a,b)_v$. It is immediate from (**??**) to deduce that this is equal to 1 almost everywhere, and that the global reciprocity law implies:

**Theorem 4.1.** $\prod_v (a,b)_v = 1$.

The Hilbert symbol is related to the $n$-th power residue symbol (i.e. the Legendre symbol for $n = 2$) and the above theorem implies other familiar reciprocity laws. For details, s. Milne, Chapter VIII.

As far as quaternion algebras go, we see that the class $[D_v]$ in $\mathrm{Br}(k_v)$ has to be non-trivial at an even (finite) number of places. In the next lecture we will see that, conversely, any even set of places defines a (unique) quaternion algebra (Hasse principle).

## 24.5 The classical formulation.

Let $\mathbb{A}_{k\infty}$ denote the product of the archimedean completions of $k$ and $\mathbb{A}_k^\infty$ the restricted product of non-archimedean completions, or *finite adeles*. Similarly for $\mathbb{A}_k^\times$.

Notice that we have a natural isomorphism: $\mathbb{A}_k^{\infty\times}/\prod_{v<\infty} \mathfrak{o}_v^\times \simeq \mathcal{I}_k$, the group of fractional ideals of $k$. Hence the ideal class group:

$$C_k = k^\times \backslash \mathbb{A}_k^{\infty\times} / \prod_{v<\infty} \mathfrak{o}_v^\times.$$

If $E/k$ is a finite Galois extension and $\mathfrak{p}$ is a prime ideal of $k$ which is unramified in $E$, and if $\mathfrak{B}$ is a prime of $E$ over $\mathfrak{p}$, one writes $(\mathfrak{B}, E/k)$ for the image of $\mathfrak{p}$ in $\mathrm{Gal}(E/k)$ under the Artin map. If $E/k$ is abelian, we will also write $(\mathfrak{p}, E/k)$ since it doesn't depend on the choice of $\mathfrak{B}$.

What we want to do now is to find an ideal-theoretic way to describe the quotients $k^\times \backslash \mathbb{A}_k^\times / K$ for all open subgroups $K$ of $\mathbb{A}_k^\times$. We define:

**Definition.** A modulus $\mathfrak{m}$ is a function $\mathfrak{m} : \{\text{places of k}\} \to \mathbb{Z}_{\geqslant 0}$ which is equal to zero almost everywhere and at all complex places, and $\leqslant 1$ at real places.

We write $v|\mathfrak{m}$ if $\mathfrak{m}(v) \neq 0$. For a modulus $\mathfrak{m}$ we let $K_\mathfrak{m}$ be the open subgroup $K_\mathfrak{m} = \prod_v K_{\mathfrak{m},v}$ of $\mathbb{A}_k^\times$ with:

$$K_{\mathfrak{m},v} = \begin{cases} \mathfrak{o}_v^\times, & \text{if } v < \infty, \mathfrak{m}(v) = 0, \\ 1 + \mathfrak{p}^{\mathfrak{m}(v)}, & \text{if } v = \mathfrak{p} < \infty, \mathfrak{m}(v) > 0, \\ k_v^\times, & \text{if } v = \infty, \mathfrak{m}(v) = 0, \\ \mathbb{R}_+, & \text{if } v = \mathbb{R}, \mathfrak{m}(v) = 1. \end{cases}$$

Such subgroups form a basis for open subgroups of $\mathbb{A}_k^\times$.

Recall that the *ray class group* $C_\mathfrak{m}$ for the modulus $\mathfrak{m}$ is defined as the quotient of an ideal group $\mathcal{I}^{\mathrm{supp}(\mathfrak{m})}$ modulo a subgroup of principal ideals $\mathcal{P}_\mathfrak{m}$ defined as follows: $\mathrm{supp}(\mathfrak{m}) = \{v|\mathfrak{m}(v) \neq 0\}$, $\mathcal{I}^S \subset \mathcal{I}$ is the subgroup of fractional ideals prime to $S$, for a set of places $S$ (ignoring possible infinite places in $S$); and $\mathcal{P}_\mathfrak{m} \subset \mathcal{I}^{\mathrm{supp}(\mathfrak{m})}$ is generated by elements $a \in \mathfrak{o}_k$ which are $\equiv 1 \mod \prod_{\mathfrak{p}<\infty} \mathfrak{p}^{\mathfrak{m}(\mathfrak{p})}$ and such that $a_v > 0$ if $v$ is a real place with $\mathfrak{m}(v) > 0$.

**Proposition 5.1.** $C_\mathfrak{m} = k^\times \backslash \mathbb{A}_k^\times / K_\mathfrak{m}$.

Now the formulations of class field theory in terms of ideals are immediate. I state only the reciprocity law and leave the rest as an exercise:

**Theorem 5.2.** *Let $E/k$ be a finite abelian extension, and let $S$ be the set of primes of $k$ ramified in $E$. We define the Artin map: $\mathcal{I}^S \to \mathrm{Gal}(E/k)$ by the properties stated in the local reciprocity law. Then this map admits a modulus $\mathfrak{m}$ with $\mathrm{supp}(\mathfrak{m}) = S$; in other words, there is such an $\mathfrak{m}$ with $\mathcal{P}_\mathfrak{m}$ in the kernel of this map. More precisely, one has an isomorphism:*

$$\mathcal{I}_k^{\mathrm{supp}(\mathfrak{m})} / \mathcal{P}_\mathfrak{m} N_k^E \mathcal{I}_E^{\mathrm{supp}(\mathfrak{m})} \simeq \mathrm{Gal}(E/k).$$

The smallest possible $\mathfrak{m}$ is called the *conductor* of $E/k$.

## 24.6 Chebotarev density.

We can apply class field theory to determine the density of primes that split in a Galois extension and answer other similar questions. The application requires the following deep theorem, whose proof (the one I know, at least) uses analytic properties of Dirichlet $L$-functions, s. Lang's *Algebraic Number Theory*, Theorem XV.6 (which is slightly stronger than what we are formulating).

**Theorem 6.1.** *Let $k$ be a number field, $P$ the set of its finite primes and let $P \to \mathbb{A}_k^\times$ be any map which sends a prime $\mathfrak{p}$ to an idele which is trivial at every place except $\mathfrak{p}$ and equal to a uniformizing element at $\mathfrak{p}$. Then (taking into account the natural filtration of $P$ by norms of its elements) the image of $P$ is equidistributed with respect to Haar measure on $\widehat{k^\times \backslash \mathbb{A}_k^\times}$. In other words, the push-forward of the uniform probability measure on $\{\mathfrak{p} | N_{\mathbb{Q}}^k(\mathfrak{p}) < M\}$ converges weak-star to uniform probability measure on $\widehat{k^\times \backslash \mathbb{A}_k^\times}$ as $M \to \infty$.*

If we combine this with class field theory, we get a weak version of Chebotarev density:

**Theorem 6.2.** *Let $E/k$ be an abelian Galois extension. The set of $\mathfrak{p} \in P$ which split completely in $E$ (equivalently, the preimage of $1 \in \mathrm{Gal}(E/k)$ under the Artin map on the set of unramified primes) has density $\frac{1}{(E:k)}$.*

A trick reduces the general case to this in order to get:

**Theorem 6.3** (Chebotarev)**.** *Let $E/k$ be a finite Galois extension, $\sigma \in \mathrm{Gal}(E/k)$ and let $C$ denote the conjugacy class of $\sigma$. The set of $\mathfrak{p} \in P$ which are unramified in $E$ and such that there exists a prime $\mathfrak{B}|\mathfrak{p}$ which maps to $\sigma$ under the Artin map has density equal to: $\frac{|C|}{(E:k)}$.*

## 24.7 The dual formulation; Weil groups; Dirichlet characters.

The fact that class field theory investigates $\mathrm{Gal}(\bar{k}/k)^{\mathrm{ab}}$ suggests that it is really a statement about *characters* (1-dimensional representations) of $\mathrm{Gal}(\bar{k}/k)$. This leads us to dualize the isomorphisms that we previously saw; the Langlands conjectures then generalize this dual formulation to include higher-dimensional representations.

Dualizing is trivial modulo the observation ("no small subgroups" argument) that a Lie group does not have arbitrarily small subgroups; therefore, if $\phi : A \to B$ is a homomorphism with $A$ locally compact totally disconnected and $B$ a Lie group then $\ker \phi$ contains an open-compact subgroup of $A$.

Let $k$ be a local field. From dualizing (**??**) we get:

$$\{\text{Complex characters of } \mathrm{Gal}(\bar{k}/k)\} \leftrightarrow \{\text{Complex characters of } \widehat{k^\times}\}.$$

In fact, we can do better than that, since we have more information about the isomorphism (**??**). Suppose $k$ is non-archimedean. Let $\mathcal{W}_k$ be the set of elements of $\mathrm{Gal}(\bar{k}/k)$ which, modulo inertia, are equal to a power of Frobenius; in other words, $\mathcal{W}_k$ is the preimage of $\mathbb{Z} \subset \hat{\mathbb{Z}}$ under the homomorphisms:

$$\mathrm{Gal}(\bar{k}/k) \to \mathrm{Gal}(k^{\mathrm{ur}}/k) \simeq \hat{\mathbb{Z}}.$$

The group $\mathcal{W}_k$ is called the *Weil group* of $k$.[3] On the other side of the reciprocity law, it corresponds to $k^\times$. We now have the bijection:

$$\{\text{Homomorphisms:}\mathcal{W}_k \to \mathbb{C}^\times = \mathbb{G}_m(\mathbb{C})\} \leftrightarrow \{\text{Irreducible representations of } k^\times = \mathbb{G}_m(k)\}. \tag{24.4}$$

Of course, we have rigged the notation so that it suggests the generalization by the Langlands program: The group $\mathbb{G}_m$ on the right hand side will be replaced by a reductive group $G$ and the group $\mathbb{G}_m$ on the left hand side will be replaced by its Langlands dual ${}^L G$.

*Remarks.* 1. Since the groups in the last assertion are not compact any more, we need to clarify what kind of "characters" or "representations" we are referring to. Indeed, in the previous (compact) formulation all characters had finite image and were unitary; this is no longer the case. In fact, for the above bijection it doesn't matter which category of characters we choose – we can choose to refer to unitary characters, or to arbitrary continuous characters. Notice that continuous characters are automatically *smooth*, i.e. locally constant; this follows from the "no small subgroups" argument. In the older literature, the term "character" often meant "unitary character", and other continuous characters were called "quasi-characters". We will adopt the (more standard nowadays) convention that "character" means "continuous homomorphism into $\mathbb{C}^\times$" will add the adjective "unitary" when needed.

2. In the non-archimedean case the distance between characters of the Weil group and those of the Galois group are not so far apart. Let us describe them on the right hand side: An *unramified character* of $k^\times$ is a character which is trivial on the maximal compact subgroup $\mathfrak{o}^\times$. On the Galois side, this means that the character is trivial on the inertia subgroup. We have an absolute value: $k^\times \to \mathbb{R}_+$, and every unramified character is of the form $x \to |x|^s$ for some complex number $s$. In fact, $s$ is uniquely defined modulo $2\pi i/\log q$, where $q$ is the order of the residue field. Now, every character $\chi$ of $k$ can be written (non-uniquely) as $\chi_0 \cdot |\ |^s$, where $\chi_0$ has finite image ($\Leftrightarrow$ extends to $\widehat{k^\times}$ or, on the Galois side, to $\mathrm{Gal}(\bar{k}/k)$).

3. There is no a priori reason to consider only complex characters. In fact, while complex representations were historically used to approach problems related to modular forms, $L$-functions etc., geometric approaches make it more natural to consider $l$-adic representations (i.e. coefficients in $\overline{\mathbb{Q}_l}$, for some prime $l$) since such are the Galois representations arising from the etale cohomology of varieties. In the local non-archimedean case, if $l$ is different from the residue characteristic of $k$ then there is no significant difference between $l$-adic and complex representations – and whichever differences arise will be discussed. If, however, $l$ is equal to the residue characteristic of $k$ then there is no analog of the "no small subgroups"

[3] There is a general abstract definition of Weil groups, which you can read in the article of Tate in the *Corvallis proceedings*.

argument, the picture is much more complex and it has been the object of intensive research since Wiles' proof of Fermat's last theorem and, especially, in recent years. This is the *p-adic Langlands program* which will be beyond the scope of our course.

In the archimedean case, we let $\mathcal{W}_{\mathbb{C}} = \mathbb{C}^\times$ and $\mathcal{W}_{\mathbb{R}}$ be the group generated by $\mathbb{C}^\times$ and an element $\sigma$ such that $\sigma^2 = -1$ and $\sigma z \sigma^{-1} = \bar{z}$ for every $z \in \mathbb{C}^\times$. We notice that in every (archimedean or not) case, $\mathcal{W}_k^{\mathrm{ab}} = k^\times$, and we have a canonical map $W_k \to \mathrm{Gal}(\bar{k}/k)$ with dense image. (These are parts of the axiomatic definition of Weil groups.) In particular, the statement of (**??**) continues to hold.

If $k$ is a global field then from dualizing (**??**) we get:

$$\{\text{Complex characters of } \mathrm{Gal}(\bar{k}/k)\} \leftrightarrow \{\text{Complex characters of } \widehat{\mathbb{A}_k^\times/k^\times}\}.$$

Characters of $\mathbb{A}_k^\times/k^\times$ are called *idele class characters* or *Grössencharacters*. Notice that if a Grössencharacter is trivial on $\mathbb{A}_{k\infty}^\times$ then it can be identified with characters of the ray class group $\mathcal{I}^{\mathrm{supp}(\mathfrak{m})}/\mathcal{P}_{\mathfrak{m}}$ for some modulus $\mathfrak{m}$ (with $\infty \nmid \mathfrak{m}$). If $k = \mathbb{Q}$ and $m = \prod_p p^{\mathfrak{m}(p)}$ then the ray class group is equal to $(\mathbb{Z}/m\mathbb{Z})^\times$, and the Grössencharacter is a *Dirichlet character*.

In the global case there is again a notion of Weil group $\mathcal{W}_k$ endowed with a homomorphism with dense image: $\mathcal{W}_k \to \mathrm{Gal}(\bar{k}/k)$, and an isomorphism: $\mathcal{W}_k^{\mathrm{ab}} = \mathbb{A}_k^\times/k^\times$, the two of them being compatible in the sense of class field theory. We won't get into details at this point, we just finish this discussion of Weil groups with a couple of remarks:

*Remarks.*

1. Notice that in the global or arhimedean case not all Weil group representations factor through Galois representations, or are even close to them in the way that we discussed in the non-archimedean case. Still, representations of non-Galois type are important for automorphic forms, for instance Maass forms are of this type. We will return to this point in the discussion of the global Langlands conjectures.

2. For the Langlands generalization, it is good to think not of representations of the group $\mathbb{A}_k/k^\times$, but of *representations of the group* $\mathbb{A}_k^\times$ *which appear on the homogeneous space* $k^\times\backslash\mathbb{A}_k^\times$. We will discuss at some point what this means.

# Chapter 25

# The automorphic space.

The theory of automorphic forms is harmonic analysis on the homogeneous space $G(k)\backslash G(\mathbb{A}_k)$, where $k$ is a global field. We call this space the *automorphic space*. This term is not completely standard, but there is no other name for it. Here we study properties of this space, discuss the adelic and the classical picture, and some relevant arithmetic issues. We fix throughout a number field $k$ (of course, almost all holds for function fields as well), and all groups are linear algebraic groups defined over $k$. The letters $S, \Sigma$ will always denote finite sets of places of $k$, $\mathbb{A}_k^S$ will denote the adeles outside of $S$, i.e. the restricted product $\prod'_{v\notin S} k_v$, and $\mathbb{A}_{k,S}$ will denote the product $\prod_{v\in S} k_v$. For a variety $X$ over $S$ we will denote: $X_k := X(k)$, $X_{\mathbb{A}} := X(\mathbb{A}_k)$, $X^S := X(\mathbb{A}_k^S)$ and $X_S := X(\mathbb{A}_{k,S})$. The (finite) set of archimedean places will be denoted by $\infty$.

## 25.1 References.

- A. Borel and G. Mostow (eds.), *Algebraic groups and discontinuous subgroups.* Proceedings of Symposia in Pure and Applied Mathematics Vol. IX, available for free on the AMS website.
- V. Platonov and A. Rapinchuk, *Algebraic groups and number theory.*
- A. Borel and Harish-Chandra, *Arithmetic subgroups of algebraic groups.* Annals of Math. (2), 75 (1962), p. 485-535.
- A. Borel, *Some finiteness properties of adele groups over number fields.* Publications mathématiques de lI.H.É.S., tome 16 (1963), p. 5-30.

## 25.2 The automorphic quotient.

Let $G$ be a linear algebraic group over $k$. Since $G$ is affine (can be embedded as a closed subvariety of affine space), the subgroup $G_k$ of $G_{\mathbb{A}}$ is discrete and the

space $[G] := G_k\backslash G_{\mathbb{A}}$ is a locally compact space, homogeneous under the action of $G_{\mathbb{A}}$. It carries an invariant measure under $G_{\mathbb{A}}$.

Basic questions regarding this space: Is it compact? If not, is the volume finite? Other finiteness properties?

We may fix an integral model for $G$, i.e. the structure of an $\mathfrak{o}$-group scheme. For instance, let $G \hookrightarrow \mathrm{GL}_n$ be a closed embedding over $k$, and consider the structure induced from the standard structure of $\mathrm{GL}_n$ (the scheme-theoretic closure of $G$ over $\mathfrak{o}$). For questions of finiteness which we will encounter here, it doesn't matter which structure we fix. For finer arithmetic questions (class numbers, etc.) which will come up later, it matters.

Let us summarize the basic results, before discussing them in more detail in the following sections:

If $G = \mathbb{G}_a$ or, more generally, a unipotent group, the space $G_k\backslash G_{\mathbb{A}}$ is compact. If $G$ is a torus, then $G_k\backslash G_{\mathbb{A}}$ is compact if and only if $G$ is anisotropic. Recall that a reductive group $G$ is called *anisotropic* if it has no non-trivial split subtori. This is more generally the case:

**Theorem 2.1.** *Let $G$ be a linear algebraic group over $k$ with maximal reductive quotient $G^{\mathrm{red}}$. The automorphic space $G_k\backslash G_{\mathbb{A}}$ is compact if and only if $G^{\mathrm{red}}$ is anisotropic.*

Regarding volumes, the basic result is:

**Theorem 2.2.** *Let $G$ be as above. Then $\mathrm{Vol}(G_k\backslash G_{\mathbb{A}}) < \infty$ if and only if $\mathcal{X}_k(G)$, the $k$-character group of $G$, is trivial.*

Finally, we have the following very basic finiteness property:

**Theorem 2.3.** *Let $G$ be as above and let $K = \prod_{v<\infty} G(\mathfrak{o}_v)$. The number of double cosets:*

$$G_k\backslash G_{\mathbb{A}}/K \cdot G_\infty$$

*is finite.*

Let us give a *Leitfaden* to the proofs and their history. All of them are contained in Borel's article mentioned in the references. Before that, Mostow and Tamagawa had given a proof of the first theorem and Borel and Harish-Chandra had given a proof of the second theorem in the "classical" setting, i.e. for $G(\mathbb{Z})\backslash G(\mathbb{R})$. The proof is based on the existence of *Siegel sets*, some very explicitly described subsets of $G(\mathbb{R})$ which contain a fundamental domain for the action of $G(\mathbb{Z})$. The passage between the adelic and the classical setting is very easy once one knows the third theorem, and we will explain it. However, as I understand it, Borel had to re-do the theory of Siegel sets in the adelic setting in order to prove the third theorem.

## 25.3 The additive group

## 25.4 The multiplicative group

## 25.5 The general linear group

The discussion of the automorphic quotient for the general linear group will depend on *strong approximation for* $\mathrm{SL}_n$, which will be discussed in more detail in the next section. It states:

$\mathrm{SL}_n(k)$ *is dense in* $\mathrm{SL}_n(\mathbb{A}_k^\infty)$.

We remind that $\mathbb{A}_k^\infty$ denotes the adeles away from infinity, i.e. the finite adeles.

Notice that this statement implies, in particular, that $\mathrm{SL}_n(\mathbb{Z})$ is dense in $\mathrm{SL}_n(\widehat{\mathbb{Z}})$, i.e. the map: $\mathrm{SL}_n(\mathbb{Z}) \to \mathrm{SL}_n(\mathbb{Z}/n)$ is surjective for every $n$. Such a result is certainly not true for the multiplicative group, for instance: $\mathbb{Z}^\times$ does not surject onto $(\mathbb{Z}/5)^\times$.

**Proposition 5.1.** *Let* $\mathrm{GL}_n(\mathbb{R})^+$ *denote the connected component of the identity in the Lie group* $\mathrm{GL}_n(\mathbb{R})$*, i.e. the group of real matrices of positive determinant. It acts with a unique orbit on the quotient:*

$$\mathrm{GL}_n(\mathbb{Q})\backslash \mathrm{GL}_n(\mathbb{A}_\mathbb{Q})/\prod_{p<\infty} \mathrm{GL}_n(\mathbb{Z}_p),$$

*and the stabilizer of the point represented by* 1 *is equal to the subgroup* $\mathrm{SL}_n(\mathbb{Z})$.

*Proof.* Given an adele $(g_2, g_3, \ldots, g_\infty)$, consider first the image of its double coset under the determinant map: It is a well-defined element of $\mathbb{Q}^\times\backslash\mathbb{A}_\mathbb{Q}^\times/\prod_{p<\infty}\mathbb{Z}_p^\times$, and as we have seen $\mathbb{R}_+^\times$ acts with a unique orbit on it. Therefore, we may assume that $(g_2, g_3, \ldots, g_\infty) \in \mathrm{SL}_n(\mathbb{A}_\mathbb{Q})$. But then, by strong approximation for $\mathrm{SL}_n$, it belongs to $\mathrm{SL}_n(\mathbb{Q})\cdot\prod_{p<\infty}\mathrm{SL}_n(\mathbb{Z}_p)\cdot\mathrm{SL}_n(\mathbb{R})$, which proves the first claim.

The stabilizer of the point represented by 1 is the intersection:

$$\mathrm{GL}_n(\mathbb{Q}) \cap \prod_{p<\infty} \mathrm{GL}_n(\mathbb{Z}_p)\cdot\mathrm{GL}_n(\mathbb{R})^+ = \mathrm{SL}_n(\mathbb{Z}).$$

□

## 25.6 Weak and strong approximation.

We say that a (geometrically integral) variety $X$ over $k$ satisfies *weak approximation* if:

For every finite set of places $S$, $X(k)$ is dense in $X_S = \prod_{v\in S} X(k_v)$.

Equivalently, if:

$X(k)$ is dense in $\prod_v X(k_v)$

the product taken over all places. We say that $X$ has the property of weak approximation away from a finite set of places $\Sigma$ if this property holds with the product taken over all places outside of $\Sigma$. For instance, if $\Sigma = \infty$ and an integral model (i.e. the structure of an $\mathfrak{o}$-scheme, where $\mathfrak{o}$ is the ring of integers in $k$) is given, then weak approximation outside of $\Sigma$ means that for every finite set of finite places $S$, every integer $N$ and every set of points $(x_v \in X(k_v))_{v \in S}$ we can find $x \in X(k)$ such that $x \equiv x_v \mod \mathfrak{p}_v^N$.

We have:

**Theorem 6.1** (Kneser, Platonov)**.** *Let $G$ be semisimple simply connected or adjoint. Then $G$ satisfies weak approximation.*

There are many more examples of groups which satisfy weak approximation, for instance $\mathrm{GL}_n$. (Proof: $\mathrm{GL}_{n,S}$ is open in $\mathrm{Mat}_{n,S}$ and carries the induced topology, so since $\mathrm{Mat}_n$ satisfies weak approximation, so does $\mathrm{GL}_n$.)

We say that a variety $X$ satisfies *strong approximation* away from a finite set of places $\Sigma$ if:

$X(k)$ is dense in $X^\Sigma = X(\mathbb{A}_k^\Sigma)$.

Sometimes if $\Sigma = \infty$ we say that $X$ satisfies strong approximation without mentioning $\Sigma$. Hence, strong approximation (away from $\infty$) is a strengthening of the statement "class number = 1". Notice that the above condition is much stronger than being dense in $\prod_v(k_v)$, because the topology on the adeles is finer than the induced topology from $\prod_v(k_v)$. For instance, if $G = \mathrm{GL}_n$ and $\Sigma = \infty$ then the property reads: For every set $S$ of finite places and for all $(x_v \in k_v)_{v \in S}$ *there exist $S$-integers* in $k^\times$ (i.e. elements of $k^\times \cap \prod_{v \in S \cup \infty} k_v^\times \prod_{v \notin S \cup \infty} \mathfrak{o}_v^\times$) which approximate $(x_v)_{v \in S}$.

A slightly weaker version of the following theorem was proven by Kneser:

**Theorem 6.2** (Platonov)**.** *If $G$ is simple, simply connected and $G_\Sigma$ is not compact then $G$ satisfies strong approximation outside of $\Sigma$.*

## 25.7 Reduction theory for $\mathrm{GL}_n$ over $\mathbb{Q}$

By "reduction theory" we mean, basically, the theory of Siegel domains and its consequences. It all relies on the special case of the group $\mathrm{GL}_n$, defined over $\mathbb{Q}$. To state the basic theorem in this case, we need some definitions:

A *fundamental domain* for the action of a discrete subgroup $\Gamma$ on a locally compact group $G$ is an open subset $D$ of $G$ such that no two points of $D$ are in the same $\Gamma$-orbit, and such that $G = \cup \gamma \in \Gamma \gamma \bar{D}$. A fundamental domain is not an object easy to describe. Therefore, one works with slightly larger sets, which may contain a finite number of translates of their points.

A *fundamental set* $\Omega$ for $\Gamma \backslash G$ is a subset of $G$ such that $\Gamma \cdot \Omega = G$ and the set $\{\gamma \in \Gamma | \gamma \Omega \cap \Omega \neq \varnothing\}$ is finite. There are also stronger (more complicated but

more convenient) conditions that one can impose, see for instance Borel's article on "Reduction theory" in "Algebraic groups and discontinuous subgroups".

Now let $G = \mathrm{GL}_n$ and consider the Borel subgroup of upper triangular matrices, with its factorization $B = AN$, where $N$ is the unipotent radical and $A$ denotes the torus of diagonal matrices. Let $\Delta$ denote the set of simple positive roots of $A$ with respect to $B$.

Let $K_\infty = \mathrm{SO}_n(\mathbb{R})$, defined with respect to the definite quadratic form represented by the unit matrix, and $K = \prod_{p<\infty} \mathrm{GL}_n(\mathbb{Z}_p) \times K_\infty$, a maximal compact subgroup of $\mathrm{GL}_n(\mathbb{A}_\mathbb{Q})$ satisfying the Iwasawa decomposition: $G(\mathbb{A}_\mathbb{Q}) = B(\mathbb{A}_\mathbb{Q}) \cdot K = N(\mathbb{A}_\mathbb{Q}) \cdot A(\mathbb{A}_\mathbb{Q}) \cdot K$.

A *Siegel set* is a subset of $G(\mathbb{A}_\mathbb{Q})$ of the form: $U_1 A_t K$ where:

- $U_1$ is a compact neighborhood of the identity in $N(\mathbb{R})$;
- $A_t = \{a \in (\mathbb{R}_+^\times)^n \subset A(\mathbb{R}) | e^\alpha(a) < t \text{ for every } \alpha \in \Delta\}$

. (There is a more general notion which does not require these specific subgroups $A, N$ and $K$, of course, which will be subsumed by the general discussion of Siegel domains later.)

Then:

**Theorem 7.1.** *There exists a Siegel set $\mathfrak{S}$ which is a fundamental set for $G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q})$.*

*Proof.* `Later.` □

## 25.8 Arithmetic subgroups, Siegel sets.

Let $G$ be an algebraic group over $\mathbb{Q}$. A subgroup $\Gamma$ of $G(\mathbb{Q})$ is called *arithmetic* if there exists a faithful representation $\rho : G \to \mathrm{GL}_n$ such that the image of $\Gamma$ is commensurable to $\rho(G) \cap \mathrm{GL}_n(\mathbb{Z})$. (Two subgroups are called *commensurable* if their intersection has finite index in both.) If this condition holds for some $\rho$, then it holds for every faithful representation of $G$ over $\mathbb{Q}$.

Let $\Gamma$ be an arithmetic subgroup, we would like to describe a fundamental domain for $\Gamma\backslash G(\mathbb{R})$. A *fundamental domain* is a connected open subset $D$ of $G(\mathbb{R})$ such that no two points of $D$ are in the same $\Gamma$-orbit, and such that $G(\mathbb{R}) = \cup \gamma \in \Gamma \gamma \bar{D}$. A fundamental domain is not an object easy to describe. Therefore, one works with slightly larger sets, which may contain a finite number of translates of their points.

A *fundamental set* $\Omega$ for $\Gamma\backslash G(\mathbb{R})$ is a subset of $G(\mathbb{R})$ such that $\Gamma \cdot \Omega = G(\mathbb{R})$ and the set $\{\gamma \in \Gamma | \gamma\Omega \cap \Omega \neq \varnothing\}$ is finite. There are also stronger (more complicated but more convenient) conditions that one can impose, see for instance Borel's article on "Reduction theory" in "Algebraic groups and discontinuous subgroups".

The following discussion assumes that $G$ is reductive just to keep ourselves from saying "the reductive quotient of $G$" all the time. Assume that $G$ is not anisotropic over $\mathbb{Q}$ and let $S$ be a maximal split torus. Let $\Phi_\mathbb{Q}$ be the relative

root system with respect to $S$, and choose an ordering of $\Phi_{\mathbb{Q}}$ (denote by $\Delta_{\mathbb{Q}}$ the simple positive roots), corresponding to a minimal parabolic $P$. The parabolic $P$ can be decomposed as $P = S \cdot M \cdot N$, where $N$ is its unipotent radical, $M$ is reductive and $\mathbb{Q}$-anisotropic and $S \cap M$ is finite. Let $K$ be a maximal compact subgroup of $G(\mathbb{R})$ whose Lie algebra is orthogonal to that of $S(\mathbb{R})$ under the Killing form – we have a decomposition $G(\mathbb{R}) = P(\mathbb{R}) \cdot K = (MN)(\mathbb{R}) \cdot S(\mathbb{R}) \cdot K$.

A *Siegel set* is a subset of $G(\mathbb{R})$ of the form: $U_1 S_t K$ where $U_1$ is a compact neighbourhood of the identity in $(MN)(\mathbb{R})$ and $S_t = \{s \in S | e^{\alpha}(s) < t \text{ for every } \alpha \in \Delta_{\mathbb{Q}}\}$. The following is the basic result of *reduction theory*:

**Theorem 8.1.** *Let $G$ be a connected reductive group over $\mathbb{Q}$ and let $\Gamma$ be an arithmetic subgroup of $G$. Let $C$ be a set of representatives for the double cosets $\Gamma\backslash G(\mathbb{Q})/P(\mathbb{Q})$ (where $P$ is as above); this set is finite. There exists a Siegel set $\mathfrak{S}$ such that $\Omega := C \cdot \mathfrak{S}$ is a fundamental set for $\Gamma\backslash G(\mathbb{R})$. Conversely, for every finite subset $C \subset G(\mathbb{Q})$ such that $C \cdot \mathfrak{S}$ is a fundamental domain for some Siegel set $\mathfrak{S}$, $C$ contains a set of representatives of $\Gamma\backslash G(\mathbb{Q})/P(\mathbb{Q})$.*

The elements of the set $\Gamma\backslash G(\mathbb{Q})/P(\mathbb{Q})$ are called the *cusps* of $\Gamma\backslash G(\mathbb{R})$. The theorem on finiteness of volume of $\Gamma\backslash G(\mathbb{R})$ now follows from computing the volume of a Siegel set, which is something explicit and straightforward.

The proof of this theorem, as well as its adelic version which will be discussed below, embeds $G$ into $\mathrm{GL}_n$ in a suitable way and then analyses the situation for $\mathrm{GL}_n$ and $\mathrm{SL}_n$.

## 25.9 The classical and the adelic picture.

Let $G$ be a connected reductive group over $\mathbb{Q}$. Fix an integral structure as above, and let $K = \prod_{v<\infty} G(\mathfrak{o}_v)$. Consider the action of $G_\infty$ on $G(\mathbb{Q})\backslash G(\mathbb{A})/K$. By the finiteness theorem **??** there are only finitely many orbits; let $\xi_i$ be representatives for those, and let $\Gamma_i \subset G(\mathbb{R})$ be the stabilizers. Hence $G(\mathbb{Q})\backslash G(\mathbb{A})/K$ is, as a $G(\mathbb{R})$-space, equal to the disjoint union of $\Gamma_i\backslash G(\mathbb{R})$. The connection between the adelic and the classical picture is established through the following easy lemma:

**Lemma 9.1.** *The subgroups $\Gamma_i$ are arithmetic subgroups of $G(\mathbb{R})$.*

For instance, if $\xi_i = 1$ then $\Gamma_i = G(\mathbb{Z})$.

Now it is clear that if $K$ is replaced by any compact open subgroup $K'$, the number of $G(\mathbb{R})$-orbits on $G(\mathbb{Q})\backslash G(\mathbb{A})/K'$ will remain finite, and the stabilizer subgroups are arithmetic. For instance, if $m = \prod_{v \in S} p_v^{m(v)}$ is a finite modulus and $K' = \prod_{v<\infty} K'_v$ where $K'_v = G(\mathfrak{o}_v)$ for $v \notin S$ and $K'_v = \{g \in G(\mathfrak{o}_v | g \equiv 1 \mod p_v^{m(v)}\}$ for $v \in S$, the stabilizer of $\xi = 1$ is the *congruence subgroup* $\{\gamma \in G(\mathbb{Z}) | \gamma \equiv 1 \mod m\}$.

## 25.10 Genus and class number.

Let $G$ be a reductive algebraic group over $k$, and fix a model for $G$ over $\mathfrak{o}$, the ring of integers of $k$. (For instance, fix an embedding $G \hookrightarrow \mathrm{GL}_n$ over $k$ and take the scheme-theoretic closure of $G$ over $\mathfrak{o}$. Notice that this will not, in general, be smooth over all points of $\mathfrak{o}$.) Denote by $K$ the compact subgroup $K = \prod_{v<\infty} K_v$ of the adeles of $G$, where $K_v = G(\mathfrak{o}_v)$.

The number of double cosets $G(k)\backslash G(\mathbb{A})/K \cdot G_\infty$ is called the *class number* of the fixed integral structure on $G$. We have seen that in the case of $G = \mathbb{G}_m$ it coincides with the class number of the field. In general, there is a more conceptual way to understand the class number: Let $G \hookrightarrow \mathrm{GL}_n$ over $k$. The integral structure on $\mathrm{GL}_n$ (and hence the induced integral structure on $G$) is induced by the choice of a *lattice* $L$ in $k^n$, i.e. a finitely-generated $\mathfrak{o}$-submodule which spans the whole space. Each such lattice gives rise to lattices $L_v := L \otimes_{\mathfrak{o}} \mathfrak{o}_v$ in $k_v^n$, and the group $G(\mathbb{A})$ acts on such collections $(L_v)_{v<\infty}$ of lattices with stabilizer $K \cdot G_\infty$. (The infinity component acts, by definition, trivially.) The orbit of $M$ under $G(\mathbb{A})$ is called the *genus* of $M$, and the number of $G(k)$-orbits in the genus is the class number.

In specific cases, there are even more conceptual explanations: Let $Q$ be a non-degenerate quadratic form in $n$ variables with integer coefficients, and let $G = O_Q$ be the orthogonal group of $Q$, considered as a subgroup of $\mathrm{GL}_n$. The Hasse-Minkowski theorem says that if $Q' \sim Q$ over $k_v$ for every $v$, then $Q' \sim Q$ over $k$. We want to ask the same question over the integers: Two integral quadratic forms $Q, Q'$ will be called *integrally equivalent* if there exists $g \in \mathrm{GL}_n(\mathfrak{o})$ with $Q \cdot g = G'$. (We denote the action of $\mathrm{GL}_n$ on the right, instead of identifying the quadratic forms with matrices and writing ${}^t g^{-1} Q g$.) We define the *genus* of $Q$ to consist of those integral quadratic forms $Q'$ such that $Q' \sim Q$ under the $\mathrm{GL}_n(\mathfrak{o}_v)$-action, for every finite $v$, and under the $\mathrm{GL}_n(k_v)$-action for infinite $k$. Then we have:

**Lemma 10.1.** *The set of integral equivalence classes in the genus of $Q$ can be naturally identified with the double coset space: $G(k)\backslash G(\mathbb{A})/K \cdot G_\infty$. Hence, the class number of $Q$ (more correctly, of the genus of $Q$) is equal to the class number of $G$, as defined previously.*

Let $Q'$ be in the genus of $Q$. In particular, $Q'$ is $k_v$-equivalent to $Q$ at every place, hence by the Hasse-Minkowski theorem $Q' = Q \cdot g$ for some $g \in \mathrm{GL}_n(k)$. This $g$ is uniquely defined modulo $G(k)$, i.e. we get a canonical element in $G(k)\backslash \mathrm{GL}_n(k)$. Now, by the fact that they are locally integrally equivalent, there exists an element $(x_v)_v \in \prod_{v<\infty} \mathrm{GL}_n(\mathfrak{o}_v) \cdot \prod_{v\in\infty} \mathrm{GL}_n(k_v)$. such that $Q' \cdot (x_v)_v = Q$, or equivalently: $g \cdot (x_v^{-1})_v \in G(\mathbb{A})$. This element is determined uniquely modulo $K \cdot G_\infty$. Hence, we have constructed a map from the genus of $Q$ to the set $G(k)\backslash G(\mathbb{A})/K \cdot G_\infty$. One now checks that this map is a bijection of the latter with the set of integral equivalence classes in the genus.

# Chapter 26

# Tamagawa numbers.

## 26.1 References:

- http://www.math.u-psud.fr/colliot/ LECTURES ON LINEAR ALGEBRAIC GROUPS BEIJING LECTURES, MORNING SIDE CENTRE, APRIL 2007 JEAN-LOUIS COLLIOT-THÉLÈNE

## 26.2 Differential forms and measures.

For what follows, we will need to endow our $p$-adic and Lie groups (and homogeneous spaces thereof) with invariant measures. It is usually best to obtain these measures from geometric objects, namely top-degree differential forms. Let $k$ be a locally compact field and let $X$ be a non-singular variety over $k$. Fix a Haar measure on $k$. For $\mathbb{R}, \mathbb{C}, \mathbb{Q}_p$ etc. we fix the standard measures (which for $\mathbb{Q}_p$ means that the measure of $\mathbb{Z}_p$ is 1). Let $\omega$ be a top-degree differential form on $X$. Then $\omega$ gives rise to a natural positive measure on $X(k)$, to be denoted by $|\omega|$. It is easy to understand what it is in local coordinates: if $X = \mathbb{A}^n$ with coordinates $x_1, \ldots, x_n$ then the differential form $dx_1 \wedge \cdots \wedge dx_n$ corresponds to the fixed measure on $k^n$. One can see that this description of the measure on $X$ does not depend on the choice of coordinates. For more details, see Weil's "Adeles and algebraic groups".

For later use, we look here at the relation between left and right Haar measures: Reductive groups carry top-degree differential forms which are left- and right- invariant, and hence their $k$-points are unimodular (i.e. left Haar measure in also right Haar measure). Unipotent groups are also unimodular. If $G = M \ltimes N$ is an algebraic group with $M$ reductive and $N$ unipotent, and if $\omega_N$ is a top-degree invariant form on $N$, then there exists a character $\mathfrak{d} : M \to \mathbb{G}_m$ which describes the action of $M$ on $\omega_N$. More precisely, if $c_m : u \mapsto mum^{-1}$ denotes the left conjugation action of $M$ on $N$, then $c_m^*(\omega_N) = \mathfrak{d}(m)\omega_N$. Let $\omega_M$ denote an invariant differential form on $M$. We have the natural projection: $p_M : G \to M$, and two different maps: $p_N^1, p_N^2 : G \to N$, such that an

element $g \in G$ is written as $p_M(g)p_N^1(g)$ or $p_N^2(g)p_M(g)$. Then one sees easily that $\omega_l := p_M^*(\omega_M) \wedge p_N^{1*}(\omega_N)$ (resp. $omega_r := p_M^*(\omega_M) \wedge p_N^{2*}(\omega_N)$) is a left (resp. right) invariant top-form on $G$. Therefore $|\omega_l|$ and $|\omega_r|$ will be left and right Haar measures, respectively. The character $\mathfrak{d}$ is the quotient between these two differential forms:

$$\omega_r = \mathfrak{d} \cdot \omega_l.$$

It is called the *modular character* of $G$.

## 26.3 Global measures

Let $k$ be a global field, and $G$ a semisimple group over $k$. Is there a natural invariant measure on the automorphic space $G(k)\backslash G(\mathbb{A})$? Since $G(k)$ acts properly discontinuouly on $G(\mathbb{A})$, this is the same as asking if there is a natural invariant measure on $G(\mathbb{A})$. Let $\omega$ be a *$k$-rational* top-degree invariant differential form on $G$. It defines measures $|\omega_v|$ on $G_v$, for every $v$. We would like to define $\mu = \prod_v |\omega_v|$ as a measure on $G(\mathbb{A})$. The problem is that this product will not converge, but let us ignore this for a moment and assume that $\mu$ is well-defined. Let $\omega'$ be another $k$-rational invariant differential form, and $\mu'$ the corresponding measure. Then we have $\omega' = a\omega$ for some $a \in k^\times$, and $\mu' = \prod_v |a|_v \mu$. By the product formula, $\prod_v |a|_v = 1$. Therefore *the measures $\mu' = \mu$ coincide!* The fact that there is an invariant measure $\mu$ on $G(\mathbb{A})$ (and, correspondingly, an invariant measure $\dot{\mu}$ on $G(k)\backslash G(\mathbb{A})$) which does not depend on any choices is amazing! This measure is called the *Tamagawa measure*.

It is now natural to ask: What is the Tamagawa measure of $G(k)\backslash G(\mathbb{A})$? The answer turns out to be something very close to 'one' – more precisely, it is the quotient of the order of the fundamental group of $G$ by the order of its Shafarevich group (=the kernel of $H^1(k, G) \to \sum_v H^1(k_v, G)$, i.e. it measures the failure of the Hasse principle), and in particular is equal to one for simply connected semisimple groups. This deceptively simple statement has taken an incredible amount of very hard work to prove (mainly by Kottwitz), and it underlies very classical and important results in arithmetic, such as the mass formula of Siegel which we will discuss below.

There remains to explain how to understand the non-convergent product: Let $K = \prod_{v<\infty} K_v$ be an open-compact subgroup of the finite adeles of $G$. We need to show how to make sense of $\prod_v |\omega_v|(K_v)$. It turns out that for almost every $v$ the factor $|\omega_v|(K_v)$ is equal to a special value $L_{G,v}(s_G)$ of the Euler factor at $v$ of a certain $L$-function like the Riemann zeta function. We have not discussed $L$-functions yet, so I'm very rapidly writing here that they are functions of a complex variable $s$ defined for $\Re s > 1$ by a convergent Euler product $\prod_v L_v(s)$, which admits meromorphic continuation. The way to make sense of the infinite product of measures, then, is to substitute $|\omega_v|$ by $\frac{|\omega_v|}{L_{G,v}(s_G)}$ – which will make the global product convergent, and in fact finite – and then multiply the product by $L_G(s_G)$ – the special value of the meromorphically continued $L$-function at $s_G$. This procedure would be problematic if $L_G$ had a

pole at $s_G$, which, however, does not happen for semisimple groups.

## 26.4 The Tamagawa measure for reductive groups

(Take everything with a grain of salt in positive characteristic; I don't understand Picard groups of algebraic groups in that case, and need to check several statements.)

## 26.5 The Tamagawa number of a reductive group

Weil conjectured that a simply connected semisimple group has Tamagawa number equal to 1. This was proven by Langlands for Chevalley groups, Lai for quasisplit groups, and Kottwitz, using the trace formula, in the general case. The general formula, including tori (a case treated by Ono), is:

**Theorem 5.1.** *The Tamagawa number of a linear algebraic group $G$ is:*

$$\tau(G) = \frac{|\operatorname{Pic}(G)_{\text{tors}}|}{|Ш(G)|}.$$

Notice that the Picard group of a linear algebraic group (we explain it below) is always finite in characteristic zero – when discussing more general cases, including the Birch and Swinnerton-Dyer conjecture, we will replace it by its torsion subgroup. $Ш(G)$ is the kernel of the map of Galois cohomology groups:

$$1 \to Ш(G) \to H^1(k, G) \to \bigoplus_v H^1(k_v, G);$$

in other words, it is the set of isomorphism classes of locally trivial $G$-torsors over $k$.

## 26.6 Picard groups of algebraic groups

Picard groups were discussed in §**??**. They were defined in terms of line bundles in the Zariski topology, but there is a similar definition in the etale topology. It turns out that the two notions coincide, see Milne, Etale cohomology, Remark 11.2 and Theorem 11.4. This can be seen as a generalization of Hilbert's theorem 90: if $X = \operatorname{spec} k$, then a Zariski line bundle is obviously trivial, but an etale line bundle is also trivial.

A line bundle is "the same" as a $\mathbb{G}_m$-bundle. Thus, by general nonsense:

$$\operatorname{Pic}(X) \simeq H^1(X, \mathbb{G}_m) := H^1(X, \mathcal{O}_X^\times),$$

or, equivalently (by the above-mentioned result), we can use the etale cohomology group $H^1_{\text{et}}(X, \mathbb{G}_m)$.

The reason for bringing in etale cohomology, is to prove the following isomorphism for torsion:

**Proposition 6.1.** *Let $k$ be an algebraically closed field and $X/k$ a proper variety. Then there is an isomorphism:*

$$\mathrm{Pic}(X)[n] \simeq \mathrm{Hom}(\pi_{1,\mathrm{et}}(X), \mathbb{Z}).$$

Here $\pi_{1,\mathrm{et}}$ is the etale fundamental group of $X$.

*Proof.* We have:

$$H^1_{\mathrm{et}}(X, \mathbb{Z}/n) = \mathrm{Hom}(\pi_1(X), \mathbb{Z}/n).$$

Furthermore, from the long exact cohomology sequence associated to:

$$0 \to \mu_n \to \mathbb{G}_m \to \mathbb{G}_m \to 1$$

we obtain an isomorphism:

$$H^1_{\mathrm{et}}(X, \mu_n) \simeq H^1_{\mathrm{et}}(X, \mathbb{G}_m)[n].$$

Here we used that $k$ is algebraically closed and that $X$ is proper over $k$; this ensures that $n$-th power map is surjective on global invertible sections:

$$H^0(X, \mathcal{O}_X^\times) \simeq k^\times \overset{n}{\twoheadrightarrow} k^\times \simeq H^0(X, \mathcal{O}_X^\times).$$

Finally, since $k$ is algebraically closed, we have $\mu_n \simeq \mathbb{Z}/n$. □

Now let us discuss covers of algebraic groups. The following is true, see here.

**Proposition 6.2.** *Every finite etale cover of an algebraic group, in characteristic zero, can be given the structure of an algebraic group if the fiber over the identity has a point.*

This statement is well-known and easy to prove for topological covers of Lie groups. It does not hold in positive characteristic. This invites the definition of *algebraic fundamental group*, which will parametrize central isogenies (which may not be etale in positive characteristic!). In any case, I think that for the torsion of the Picard group the statement of Proposition **??** holds whether we use the etale or the algebraic fundamental group. In particular:

**Corollary 6.3.** $\mathrm{Pic}(G)$ *is trivial if $G$ is semisimple simply connected in the sense of algebraic groups (i.e.: its coweight lattice is generated by coroots).*

*Remark.* If $G$ is a torus, there is an isomorphism: $\mathrm{Pic}(T) \simeq H^1(\mathrm{Gal}(\bar{k}/k), \mathcal{X}^*_{\bar{k}}(T))$. See Colliot-Thélène, Proposition 4.23. This shows that the calculation of Tamagawa numbers of tori by Ono matches the aforementioned result.

*Remark.* For the computation of the Tamagawa number of an arbitrary reductive algebraic group from that of a simply connected one, see Sansuc, "Groupe de Brauer et arithmétique des groupes algébriques linéaires sur un corps de nombres."

## 26.7 The work of Siegel on quadratic forms

TO BE ADDED.

# Part IV

# Automorphic representations

This part is very incomplete and by no means justifies the word ‘‘automorphic’’ in the title.

# Chapter 27

# Basic representation theory of real and $p$-adic groups.

## 27.1 References

- Goodman and Wallach, *Representations and invariants of the classical groups.*
- W. Schmid, *Geometric Methods in Representation Theory.* Notes from a mini-course, available at:
  `www.math.harvard.edu/∼schmid/articles/brussels_1_30_04.pdf`
- M. Welleda Baldoni, *General Representation Theory of Real Reductive Lie Groups*, in the *Edinburgh Proceedings* (PSPUM 61).
- Knapp, *Lie groups beyond an introduction.*
- T. Bröcker and T. Tom Dieck, *Representations of compact Lie groups.*
- Varadarajan, *Lie groups, Lie algebras and their representations.*
- D. Bump, *Lie groups.*
- W. Casselman, *Introduction to the theory of admissible representations of p-adic reductive groups.* Available online at:
  `http://www.math.ubc.ca/ cass/research/p-adic-book.dvi`.
- J. Bernstein, *Lectures on representations of reductive p-adic groups.* Notes by Karl Rumelhart. Available online at:
  `http://www.math.uchicago.edu/∼mitya/langlands/Bernstein/Bernstein93new.dvi`.
- J. Bernstein, *Le "centre" de Bernstein.* Edited by P. Deligne. Travaux en Cours, Representations of reductive groups over a local field, 1–32, Hermann, Paris, 1984.

- P. Cartier, *Representations of p-adic groups: A survey.* In the Corvallis proceedings, Vol 1. Available online at: `http://www.ams.org/online_bks/pspum331/`.
- C. Moeglin, *Representations of* $\mathrm{GL}(n, F)$ *in the nonarchimedean case.* In: T. N. Bailey and A. W. Knapp, *Representation Theory and Automorphic forms*, Proceedings of Symposia in Pure and Applied Mathematics, AMS 1997.

## 27.2 Continuous representations.

For the first part of the notes for this lecture, I'm just collecting the very basic facts of representations of topological and Lie groups, many of which you can find in more detail in the (concise and highly recommended) notes of Schmid or the article of Welleda Baldoni.

Let $G$ be a locally compact topological group, $V$ a complete, Hausdorff, locally convex topological vector space (over $\mathbb{C}$). A *(continuous) representation* of $G$ on $V$ is a representation of $G$ on $V$ such that the action map: $G \times V \to V$ is continuous.

A representation is called *irreducible* if $V$ does not have any proper closed invariant subspaces.

*Remark.* Let $V$ be a Banach space, and $B(V)$ the Banach space of bounded endomorphisms of $V$. Then one could require that $G \to B(V)$ is continuous. However, this is too strong to ask. Exercise: The regular representation of $\mathbb{R}$ on $L^p(\mathbb{R})$ $(p < \infty)$ is not continuous in this sense.

The assumptions on $G$ and $V$ imply that the following make sense:

- (Left and right) Haar measures on $G$ (because of local compactness of $G$).
- (Lebesgue) integration of functions with values in $V$ (because of completeness and local convexity of $V$).
- If $M$ is a manifold, differentiable functions $M \to V$ (e.g. if $M = \mathbb{R}$ then $f'(0) = \lim_{h \to 0} \frac{f(h)-f(0)}{h}$).

In particular, for every $f \in C_c(G)$ (continuous, compactly supported functions on $G$), $v \in V$ the integral $\pi(f)v = \int_G f(g)\pi(g)v dg$ (where $dg$ is a Haar measure – assume here that $G$ is unimodular, i.e. left Haar = right Haar to avoid complications) is defined and convergent, which gives rise to a representation of the convolution algebra $C_c(G)$ on $V$. (Convolution: $(f_1 * f_2)(g) = \int_G f_1(x) f_2(x^{-1}g) dx$.) In fact, $C_c(G)$ is a $*$-algebra, i.e. equipped with a sesquilinear antiinvolution, which is: $f \mapsto f^*(g) := \overline{f(g^{-1})}$, and the representation is a $*$-representation.

*Remark.* It is more natural to think of convolution algebras of *measures* than of *functions*, because then the action and the algebra structure does not depend on choices of Haar measures. This is the approach that we will mostly follow.

If $V$ is a Hilbert space, the representation $\pi$ is called *unitary* if $\pi(g)$ is a unitary automorphism for every $g$. Our main focus will be on representations of groups on Hilbert spaces. However, they will not always be unitary. Therefore, the norm of the space is in many cases irrelevant; we can instead speak of *Hilbertian spaces*, i.e. topological vector spaces which can be given the structure of a Hilbert space. For instance, all finite-dimensional vector spaces are Hilbertian. A representation $\pi$ of $G$ on a Hilbertian space $V$ is *unitarizable* if $V$ admits a Hilbert structure which makes $\pi$ unitary.

## 27.3 Continuous representations of compact groups.

**Theorem 3.1** (Peter-Weyl)**.** *Let $G$ be a compact group. Every representation of $G$ on a Hilbertian space $V$ is unitarizable and decomposes into a Hilbert space direct sum of irreducibles, all of which are finite-dimensional.*

We first prove a slightly stronger theorem for the special case of $V = L^2(G)$. Here we have an action of the group $G \times G$ by left and right multiplication. Given a representation $\tau$ of $G$, the *matrix coefficient* of $\tau$ is the canonical map: $\tau \otimes \tau^* \to C(G)$, where $\tau^*$ denotes the dual.

**Theorem 3.2.** *The matrix coefficients induce an isomorphism: $L^2(G) \simeq \overline{\oplus_\tau(\tau \otimes \tau^*)}$, the (orthogonal) sum ranging over all isomorphism classes of irreducible finite-dimensional representations of $G$. Moreover, the dense subspace $\oplus_\tau(\tau \otimes \tau^*)$ is also dense in $C(G)$.*

*Proof (Sketch):* Let $L$ denote the left regular representation and $R$ the right one. One first proves:

**Lemma 3.3.** *For a function $f$ on $G$: $f$ is left $G$-finite (i.e. spans a finite-dimensional subspace under $L(G)$) $\iff$ it is right $G$-finite $\iff$ it is a matrix coefficient for a finite-dimensional representation of $G$.*

Given this, the proof of the theorem relies on the fact that for $f \in C(G)$ the operator $L(f)$ is compact, and if $f = f^*$ then it is self-adjoint. Applying the spectral theorem, we get many finite-dimensional eigenspaces (corresponding to all non-zero eigenvalues) which are $R(G)$-invariant and hence, by the lemma, spaces of finite-dimensional matrix coefficients. (The rest and more detail you can find e.g. in Bump's book.) □

A representation of a group $G$ on a vector space is called *locally finite* if the $G$-span of every vector is finite dimensional.

Now we prove the Peter-Weyl theorem. (By the way, the last theorem is also called "the Peter-Weyl theorem".

*Proof (Sketch):* First, start with any Hilbert space norm, average it over $G$, this will give a Hilbert space norm with respect to which the action is unitary.

To prove the existence of invariant finite-dimensional subspaces, one acts on $V$ by an invariant finite-dimensional subspace of $C(G)$ – using the previous theorem which says that these spaces span a dense subspace of $C(G)$.

The decomposition is now a routine application of Zorn's lemma and the fact that the orthogonal complement of an invariant subspace is also invariant. □

## 27.4 Finite-dimensional representations of Lie groups.

We now discuss representations of real and complex Lie groups, which we will throughout assume *reductive* and (for simplicity) *algebraic* and *geometrically connected*.

A *smooth* representation of $G$ on a topological vector space $V$ as previously is one such that for every $v \in V$ the orbit map $g \mapsto \pi(g)v$ is smooth (i.e. $C^\infty$). It then extends to a representation of $\mathfrak{g}$. (Recall that a representation of a Lie algebra is the same thing as a representation of its universal enveloping algebra.) We will agree that for us *smooth* will also mean that $V$ is a Fréchet space – as we will note, this is the case for the smooth vectors of a Banach representation.

**Lemma 4.1.** *Let $(\pi, V)$ be a smooth representation of a Lie group $G$. Assume that $V$ is finite-dimensional. Then every closed $\mathfrak{g}$-invariant subspace is $G$-invariant (and vice versa, of course, which is obvious).*

This Lemma does not hold without the assumption that $V$ is finite-dimensional: Consider the subspace of $C^\infty(G)$ consisting of functions supported in a given closed set!

### 27.4.1 The unitarian trick

We start with an investigation of finite-dimensional representations.

**Theorem 4.2.** *Every complex semisimple Lie group $G$ admits a unique (up to $G$-conjugacy)* compact real form $U \subset G$.

**Corollary 4.3.** *Every finite-dimensional representation $(d\pi, V)$ of a complex semisimple Lie algebra $\mathfrak{g}$ is semisimple.*

*Proof.* Let $G$ be the simply connected complex semisimple group with Lie algebra $\mathfrak{g}$, then the representation $(d\pi, V)$ lifts to a representation $(\pi, V)$ of $G$. If $U \subset G$ is as in the above theorem, then $\pi|_U$ is semisimple. By Lemma **??**, this is equivalent to $d\pi|_{\mathfrak{u}}$ being semisimple. It follows that $d\pi$ is semisimple. □

This is the "*unitarian trick*" of Weyl. There are purely algebraic proofs of semisimplicity, see Goodman and Wallach.

### 27.4.2 Weights; the Cartan decomposition.

Let $(\Phi, V)$ be a root system, $(\check{\Phi}, V^*)$ its dual root system. Let $\check{R} \subset V^*$ be the lattice spanned by the co-roots. The *weight lattice* is the sublattice $\{v \in V | \langle v, r\rangle \in \mathbb{Z} \text{ for all } r \in \check{R}\}$. It (properly, in general) contains the root lattice.

Let $A$ be a torus, the *weights* of $A$ are the characters of $A$.

Let $G$ be a reductive group and $(X, \Phi, \check{X}, \check{\Phi})$ its root datum. The weights of $G$ are the weights of its maximal torus, i.e. the elements of $X$. They do not, in general, coincide with the weights of its Lie algebra (they do iff $G$ is semisimple), and therefore to distinguish them we often call them *integral weights*. If we fix a Weyl chamber $\mathcal{C}$, we call *dominant* the weights of $G$ belonging to $\mathcal{C}$.

Returning to the setting of Lie groups, we notice that if $A$ is a complex torus and $T$ its compact real form, then the weights of $A$ correspond precisely to the (necessarily unitary) characters of $T$, i.e. the Plancherel dual of $T$.

Now we discuss maximal compact subgroups of real reductive Lie groups and the Cartan involution.

**Theorem 4.4.** *Let $G$ be a non-abelian, real, reductive (algebraic) group. There is a unique up to conjugacy maximal compact subgroup $K$. Let $\mathfrak{g}$, $\mathfrak{k}$ denote the Lie algebras. There is a unique $(ad)(\mathfrak{k})$-invariant subspace $\mathfrak{p}$ of $\mathfrak{g}$ with the properties:*

1. *$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$.*
2. *$[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$; equivalently (in combination with $[\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}$), $\mathfrak{k}$ is the $+1$ eigenspace and $\mathfrak{p}$ the $-1$ eigenspace of an involution $\theta$ of $\mathfrak{g}$ – the* Cartan involution*.*
3. *Every element of $\mathfrak{p}$ is diagonalizable over $\mathbb{R}$.*

*The map $K \times \mathfrak{p} \to G$: $(k, X) \mapsto k \cdot \exp(X)$ is a diffeomorphism of manifolds.*

The Cartan involution lifts to an involution of $G$. If $\mathfrak{g}$ is semisimple and $B$ is the Killing form, it can also be characterized as follows:

**Theorem 4.5.** *The Cartan involution is the unique – up to conjugacy – involution on $\mathfrak{g}$ such that the symmetric bilinear form $B_\theta(X, Y) := -B(X, \theta Y)$ is positive definite.*

Moreover, notice that by the last statement of Theorem **??**, the inclusion $K \hookrightarrow G$ is a strong deformation retract and, in particular, induces isomorphisms of homotopy and homology groups.

We notice that the compact form $U$ of $G_{\mathbb{C}}$, which we talked about before, is that with Lie algebra $\mathfrak{k} \oplus i\mathfrak{p}$, and $K = U \cap G$ as subgroups of $G_{\mathbb{C}}$.

### 27.4.3 Highest weight theory

Let now $G$ be a complex (algebraic) Lie group and $U$ its compact real form (Theorem **??**).

**Theorem 4.6.** *For each (integral) dominant weight $\lambda$ there is a unique (up to isomorphism) irreducible finite-dimensional representation $V_\lambda$ of $U$ with highest weight $\lambda$, and all irreducible finite-dimensional representations are of this form.*

*Weyl's proof.* Every finite-dimensional representation $\pi$ must have a highest weight. Notice that the character of $\pi$ – which is a $W$-invariant function on $T$[1] – is equal to the sum of weights appearing, with the multiplicity that they appear. One writes down an "explicit" $\mathbb{Z}$-basis for such functions on $T$[2] and then uses the orthogonality relations for characters $\chi_\pi$, together with the Weyl character formula, to show that each character should be precisely equal to one element in this basis. Since the characters span the space of class functions (this follows from the Peter-Weyl theorem), all elements of the basis should be characters of some reprensentation. □

Corollaries of the character computation are:

1. Each irreducible representation has a *unique* highest weight $\lambda$, which appears with multiplicity one.
2. The other weights appearing are (precisely) those $\mu$ such that ${}^w\mu < \lambda$, where ${}^w\mu$ is a translate of $\mu$ lying in the dominant Weyl chamber.
3. The characters are regular (polynomial) functions on $U$.

As a corollary of the third remark, all finite-dimensional representations of $U$ are algebraic. This, together with Corollary **??**, implies:

> There is an equivalence of categories between the category of algebraic representations of $G$ and locally finite (or Hilbertian, by Peter-Weyl) representations of $U$.

Notice that algebraic (by definition, those whose matrix coefficients are regular functions on $G$) representations are automatically locally finite, because matrix coefficients embed the space spanned by any vector into $\mathbb{C}[G]$, and the action of $G$ on the latter is locally finite.

For this reason, Theorem **??** holds for algebraic representations of $G$, as well. Of course there are also algebraic proofs of the theorem, see Goodman and Wallach.

## 27.5 Infinite-dimensional representations of Lie groups.

Here is a problem which we are facing with infinite dimensional representations: Which of them should be considered 'equivalent'? For instance, all $L^p(\mathbb{R})$

[1] Each conjugacy class in $U$ is represented by an element of $T$, unique modulo the $W$-action.

[2] It's actually not as simple as that: One works with anti-symmetric instead of symmetric functions first.

($p < \infty$) should be thought of as 'equivalent' representations of $\mathbb{R}$, though they are non-isomorphic as Banach spaces. There is a good answer to this question (of algebraic nature, somehow) in the case that the representations under consideration are *admissible*.

### 27.5.1 Admissibility

We return to the general setting of a topological vector space $V$ as above, and a representation of a Lie group $G$ on $V$. We let $K$ be the maximal compact subgroup of $G$. It is natural to first restrict any representation $\pi$ of $G$ to $K$, since the picture for compact groups is relatively simple.

**Definition.** A representation $(\pi, V)$ as above is *admissible* if for every finite-dimensional irreducible representation $\tau$ of $K$ we have $\dim \mathrm{Hom}_K(\tau, \pi) < \infty$.

It is not true that every irreducible representation of $G$ (not even on a Banach space) is admissible. However, Harish-Chandra's admissibility theorem states:

**Theorem 5.1.** *Every irreducible unitary representation of $G$ is admissible.*

### 27.5.2 Smooth and $K$-finite vectors.

Now, let $(\pi, V)$ be a representation of $G$. We let $V^\infty$ denote the space of smooth vectors in $V$, and $V_{K\text{-fin}}$ the space of $K$-finite vectors. Then:

**Theorem 5.2.** *If $(\pi, V)$ is admissible then $V_{K\text{-fin}}$ is contained in $V^\infty$ and is dense in $V$.*

This is proved by considering the action of the algebras $C_c^\infty(G)$ and $C(K)_{K\text{-fin}}$ and knowledge of density and regularity properties of these subalgebras of $C_c(G)$ and $C(K)$.

As a corollary, by applying the above to the case $G = K$, we get the below theorem which for Hilbertian representations we know already from compact-group generalities:

**Corollary 5.3.** *Every irreducible representation of $K$ is finite dimensional.*

### 27.5.3 Harish-Chandra modules.

Given a representation $(\pi, V)$ of $G$ we can now restrict our attention to the dense subspace $V^\infty$ which is a $\mathfrak{g}$-module. There is a problem here: It may have many $\mathfrak{g}$-invariant subspaces, even if $V$ is irreducible.

*Remark.* The space $V^\infty$ is called the *Gårding space*, and it can be given a natural topology making it a topological vector space as in §**??**. If $V$ is a Banach space, then this topology makes $V^\infty$ into a Fréchet space. This still doesn't solve our problem: closed $\mathfrak{g}$-invariant subspaces of $V^\infty$ will not necessarily have $G$-invariant closures in $V$.

The problem is solved via the following observation:

**Lemma 5.4.** *The space $V_{K\text{-fin}}$ is $\mathfrak{g}$-invariant.*

*Proof.* Indeed, for every $v \in V_{K\text{-fin}}$ the space $d\pi(\mathfrak{g})\pi(K)v$ is finite-dimensional since $\mathfrak{g}$, and for every $X \in \mathfrak{g}, k \in K$ we have: $\pi(k)d\pi(X)v = d\pi(\mathbb{A}(k)X)\pi(k)v \in d\pi(\mathfrak{g})\pi(K)v$. □

**Definition.** A $(\mathfrak{g}, K)$-*module* is a triple $(M, \pi, d\rho)$ where $M$ is a (complex) vector space (no topology!), $\pi$ is a *locally finite* representation of $K$ on $V$, $d\rho$ is a representation of $\mathfrak{g}$ and $\pi, d\rho$ are compatible, in the sense that $d\pi = d\rho$ and $\pi(k)d\rho(X)\pi(k^{-1}) = d\rho(\mathrm{Ad}(k)X)$. A $\mathfrak{g}$-finitely generated and admissible $(\mathfrak{g}, K)$-module is called a *Harish-Chandra* module.

We also define *admissibility* for a $(\mathfrak{g}, K)$-module in the same way that we did for a $G$-representation. The notion of a $(\mathfrak{g}, K)$-module solves the problem of invariant subspaces – this rests upon the fact that for admissible representations the $K$-finite vectors are analytic. Moreover, every Harish-Chandre module can be globalized, i.e. it comes from a smooth representation of $G$. All this is contained in the Casselman-Wallach theorem, which states:

**Theorem 5.5.** *The functor $V \mapsto V_{K\text{-fin}}$ is an equivalence of categories between admissible, finitely generated smooth representations of $G$ and Harish-Chandra modules.*

Harish-Chandra had defined two representations $V_1, V_2$ of $G$ to be *equivalent* if $V_1^\infty \simeq V_2^\infty$ as smooth representations. On the other hand, we say that $V_1$ and $V_2$ are *infinitesimally equivalent* if $V_{1K\text{-fin}} = V_{2K\text{-fin}}$ as $(\mathfrak{g}, K)$-modules. The theorem of Casselman and Wallach tells us that these two notions coincide for finitely generated, admissible representations.

## 27.6 Representations of $l$-groups: the Hecke algebra.

The situation is significantly simpler from the analytic point of view for $p$-adic groups. (On the other hand, it is much more complicated from the algebraic / representation-theoretic point of view: classifying irreducible representations is much harder.)

Let $G$ be a locally compact, totally disconnected group. (Such a group is called an $l$-group; a topological space which is Hausdorff, locally compact and totally disconnected is called an $l$-space.) The identity has a topological basis consisting of open, compact subgroups (exercise!). A representation of $G$ on a vector space $V$ (no topology[3]) is called *smooth* if every $v \in V$ has an open stabilizer. We let $\mathcal{H}(G)$ denote the convolution algebra of locally constant, compactly supported measures on $G$ – it is called the *Hecke algebra* and acts on every smooth representation of $V$. It is an *idempotented algebra*: for every

[3] One can actually endow a smooth representation with a natural topology, see "Bernstein's Center", but that doesn't help much.

finite set $\{f_i\}_i \subset \mathcal{H}(G)$ there exists an idempotent $e$ (i.e. $e^2 = e$) such that $ef_i = f_i e$ for all $i$. In fact, such idempotents are the characteristic measures $e_K$ of compact open subgroups $K$. A module $V$ for $\mathcal{H}(G)$ is called *non-degenerate* or *unital* if $\mathcal{H}(G)V = V$. The following is easy:

**Theorem 6.1.** *The natural functor is an equivalence of categories between smooth $G$-representations and non-degenerate $\mathcal{H}(G)$-modules.*

Let $K$ be an open compact subgroup of $G$. We define admissibility as previously: A representation $(\pi, V)$ of $G$ is admissible if for every admissible $\tau$ of $K$ the space $\mathrm{Hom}_K(\tau, \pi)$ is finite dimensional. Notice that this notion does not depend on the choice of $K$. For smooth representations, there is also the following equivalent formulation:

> $(\pi, V)$ is admissible if for every compact open $K \subset G$ the space $V^K$ (of $K$-invariants) is finite dimensional.

Given a smooth representation $\pi$ we denote by $\pi^*$ its dual and by $\tilde{\pi}$ the space of smooth vectors in $\pi^*$. The following is straightforward:

**Lemma 6.2.** *$\pi$ is admissible if and only if $\pi \hookrightarrow \tilde{\tilde{\pi}}$ is an isomorphism.*

There are many interesting non-admissible representations, for instance (usually) the space $C_c^\infty(X)$ when $X$ is an $l$-space with a $G$-action.

Since we do not consider any topology on the space of a smooth representation, irreducibility will be defined in the algebraic way: If there are no proper, non-zero, invariant subspaces. We have Schur's lemma:

**Theorem 6.3.** *Assume that $G$ is "countable at infinity", i.e. its one-point compactification has a countable basis of neighbourhoods, i.e. $G$ is a countable union of compact subsets. Let $V$ be an irreducible smooth representation. Then* $\mathrm{End}_G V = \mathbb{C}$.

*Proof.* See Bernstein's notes, §4.2. □

For any compact open subgroup $K \subset G$, let $\mathcal{H}(G, K) = e_K * \mathcal{H}(G) * e_K =$ the convolution algebra of $K$-biinvariant measures on $G$. This is the *$K$-Hecke algebra*. On the relation between Hecke modules and irreducible smooth representations, we have:

**Theorem 6.4.** *Let $(\pi, V)$ be a smooth representation of $G$. It is irreducible if and only if for every open compact subgroup $K$ the module $V^K$ of $\mathcal{H}(G, K)$ is irreducible.*

*Two irreducible representations $(\pi_1, V_1)$ and $(\pi_2, V_2)$ are equivalent if and only if for some $K$ the $\mathcal{H}(G, K)$-modules $V_1^K, V_2^K$ are non-zero and equivalent.*

*Finally, every irreducible $\mathcal{H}(G, K)$-module is equal to $V^K$, for some irreducible representation $V$ of $G$.*

*Proof.* See Bernstein's notes, §4.2. □

Let now $G$ denote the group of $k$-points of a reductive group over $k$, a $p$-adic field. The following is a combination of theorems of Harish-Chandra, Jacquet and Howe:

**Theorem 6.5.** *For every representation $\pi$ of $G$ on a topological vector space (as in §??) the space $V^\infty$ of smooth vectors is dense. Every irreducible unitary representation of $G$ is admissible. Every irreducible smooth representation of $G$ is admissible.*

*Proof.* The first statement is obtained easily by integrating a vector over a very small subgroup. The second follows from the other two. So the third is really the essential part. This is proven first for a certain class of representations called "supercuspidal" and then by showing that all irreducible representations are obtained as subquotients of: (start with supercuspidal on a Levi subgroup) $\rightsquigarrow$ (consider it as a representation of a parabolic subgroup with this Levi) $\rightsquigarrow$ (induce to $G$). These are topics which we will discuss in a future lecture. □

# Chapter 28

# Automorphic forms and the Hecke algebra.

In this lecture we will see several "approximations" to the notion of an automorphic form and automorphic representation, and eventually give a proper definition. We will also discuss several sources of interest in the subject.

## 28.1 The case of $SL_2$ – classical approach.

Let $\mathbb{H} = \{z \in \mathbb{C} | \Im(z) > 0\}$, the complex upper-half-plane. The group $G = \mathrm{SL}_2(\mathbb{R})$ acts on the left by fractional linear transformations: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}$. The action is transitive, and the stabilizer of $z = i$ is the compact form of $\mathrm{SO}_2$, to be denoted by $K_\infty$. Hence we can write: $\mathbb{H} = G/K_\infty$.

*Remark.* Notice that the center of $\mathrm{SL}_2(\mathbb{R})$ acts trivially, therefore we should really think of the action of $\mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$. However, this is not the same as $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{PGL}_2(\mathbb{R})$!

Any discrete subgroup $\Gamma$ of $G$ acts properly discontinuously on $\mathbb{H}$ (= for any $x, y \in \mathbb{H}$ there exist neighbourhoods $U_x, U_y$ such that $\{\gamma \in \Gamma | \gamma U_x \cap U_y \neq \varnothing\}$ is finite. Therefore, the quotient $\Gamma\backslash\mathbb{H}$ is well-defined as a locally compact, Hausdorff topological space. In fact, more is true: If $\Gamma$ (or $\Gamma/\{\pm 1\}$) acts freely (which in this case is equivalent to having trivial stabilizers at all points) then there is a canonical complex structure for the quotient.

If $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ then it doesn't act freely (the stabilizers of $z = i, e^{\frac{2\pi i}{6}}$ and their translates are non-trivial), but the quotient can still be endowed with a complex structure. We notice that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the elements $\begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$ and $\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$.

We start 'approximating' the notion of an automorphic form, in the case of $\mathrm{SL}_2$. (In this case there are two distinct families of automorphic forms: Maaß forms and holomorphic modular forms. We will see later where this distinction comes from. In the beginning we will take an approach leading us to Maaß forms.)

> **0th approximation:** Automorphic forms on $\Gamma\backslash\mathbb{H}$ are functions on $\Gamma\backslash\mathbb{H}$.

This is too crude; we have to specify what kind of functions. The idea, eventually, will be that we have a space $X$ together with the action of a group $G$ and hence a representation of $G$ on the space of functions on $X$. Then we single out the functions which belong to 'irreducible subspaces', as a particularly good 'basis' for functions on $X$. Let us however postpone the group-theoretic approach, and stick to a classical point of view for now.

The space $\mathbb{H}$ admits a hyperbolic metric $ds = \frac{dxdy}{y^2}$, with respect to which it is complete. The group $\mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$ is simultaneously the group of conformal transformations and the group of isometries! In particular, the metric descends to the quotient $\Gamma\backslash\mathbb{H}$. We consider the hyperbolic Laplace-Beltrami operator:[1] $\Delta = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$.

So, we pose the problem: describe the spectrum of $\Delta$ – find its eigenfunctions on $\Gamma\backslash\mathbb{H}$.

The problem is not very meaningful a priori – we need to discuss what kind of spectrum we have. Let us assume for a moment that $\Gamma\backslash\mathbb{H}$ is compact. As we have seen, this is not the case with arithmetic subgroups of $\mathrm{SL}_2$, but it is true for arithmetic subgroups of quaternion division algebras. It is also true for non-arithmetic subgroups $\Gamma$ which appear as fundamental groups of compact, genus$>1$ Riemann surfaces.

*Remarks.* 1. At this point we should explain the connection to Riemann surfaces, which also explains the interest of other areas of mathematics in the space which we are discussing: Let $X$ be a compact, connected Riemann surface of genus$>1$. Then its fundamental cover (as a complex manifold) is precisely the upper half plane $\mathbb{H}$, and hence $X = \Gamma\backslash\mathbb{H}$, where $\Gamma = \pi_1(X)$. (The group $\Gamma$ acts freely in this case.) By the way, it often happens that problems which people want to solve for general Riemann surfaces are not approachable, in which case one may be able to do better in the special case that $\Gamma$ is arithmetic, in which we have more tools. See, for instance, the area of "arithmetic quantum chaos".

2. While we are there: In contrast to $\mathrm{PSL}_2(\mathbb{R})$, the arithmeticity (or *arithmetic rigidity*) theorem of Margulis asserts that, if $G$ is a connected semisimple Lie group with trivial center, no compact factors and real

[1] Note on the sign conventions: Usual definitions of the Laplacian are as $\Delta f = \operatorname{div}\operatorname{grad} f$, and another definition using the complex of differential forms and the $d, \delta$ operators ($\Delta = d\delta + \delta d$). The latter is minus the former, and it is the convention that we follow here.

rank greater than one, then any discrete subgroup of cofinite volume is arithmetic.

Returning to the Laplacian, one can show that when the quotient $\Gamma\backslash\mathbb{H}$ is compact then $\Delta$ is a self-adjoint (unbounded) operator on $L^2(\Gamma\backslash\mathbb{H})$. Therefore, spectral theory together with the fact that the Laplacian is elliptic implies that $L^2(\Gamma\backslash\mathbb{H})$ admits an orthonormal basis of eigenvectors for $\Delta$, with the sequence of real (non-negative, in the case of the Laplacian) eigenvalues tending to infinity and that all eigenvectors will be smooth functions. Here is a better approximation to the notion of an automorphic form:

> **1st approximation:** Automorphic forms on $\Gamma\backslash\mathbb{H}$ are eigenfunctions for $\Delta$ on $\Gamma\backslash\mathbb{H}$.

In fact, we have just defined the notion of *Maaß forms*. (In the non-compact case, we would have to add some growth condition to the definition.)

## 28.2 Representation-theoretic approach.

We now want to interpret the above representation-theoretically. The first step is easy: we lift our functions to $G$. Let's agree from now on that we will consider smooth functions only – we have motivated this by observing that eigenfunctions of the Laplacian will always be smooth. Hence, instead of thinking of functions on $\Gamma\backslash\mathbb{H}$, we think of $K_\infty$-invariant functions (those are also called *spherical functions*) on $\Gamma\backslash G$. The benefit here is that we have an action of the group $G$ on this space. Therefore, depending on which category of representations we prefer to work in, we may consider the space generated by $G$-translates of these functions, which is a smooth representation of $G$, or we may choose to consider the $(\mathfrak{g}, K_\infty)$-module generated by those functions, to preserve the property of $K_\infty$-finiteness.

What does this have to do with the Laplacian which we were considering earlier? The group $G$ does not act on functions on $\Gamma\backslash\mathbb{H}$, and neither does its Lie algebra. However, the center $\mathfrak{z}(\mathfrak{g})$ of the universal enveloping algebra acts, since it commutes with the action of $K$! In the case of $\mathrm{SL}_2$, $\mathfrak{z}(\mathfrak{g})$ is generated by the Casimir element $C$, and it turns out that $C = -2\Delta$.

So, how does the requirement that our functions be eigenvectors of the Laplacian arise from representation theory? Let $(\pi, V)$ be an irreducible Harish-Chandra-module of a Lie group $G$. (Equivalently, we could talk about an irreducible admissible representation.) Then $\mathfrak{z}(\mathfrak{g})$ acts by a scalar on $\pi$ (this is Schur's lemma, which I forgot to write about, but follows easily from admissibility), i.e. there is an eigencharacter $\psi$ of $\mathfrak{z}(\mathfrak{g})$ such that $\pi(D)v = \psi(D)v$ for every $D \in \mathfrak{z}(\mathfrak{g})$. Therefore, we can strengthen approximation 1 in a natural representation-theoretic way as:

> **2nd approximation:** An automorphic representation on $\Gamma\backslash\mathbb{H}$ is an irreducible Harish-Chandra module $\pi$ of $G$, possessing a $K_\infty$-invariant vector, together with an embedding $\nu : \pi \to C^\infty(\Gamma\backslash G)$.

Of course, as we shall see, the requirement that there are spherical vectors is irrelevant, and removing it will lead us to automorphic forms other than the Maaß forms which we have been discussing so far.

The above is, in fact, an almost precise definition for an *automorphic representation* on $\Gamma\backslash G$. (Though the adelic definition which we will give in a while is a bit finer because it decomposes the $\pi$-isotypic space further with respect to the action of $G(\mathbb{A}_f)$.) The word "almost" refers to the fact that, since the categories of representations which we are considering are not semi-simple, it is not justified to embed only irreducible representations.

## 28.3 Automorphic forms: precise definition (classical).

Let $G, \Gamma$ be as above and $\nu : \pi \to C^\infty(\Gamma\backslash G)$ and embedding of an irreducible Harish-Chandra module. Functions in the image of $\nu$ satisfy the following properties:

1. They are $K_\infty$-finite.
2. They are $\mathfrak{z}(\mathfrak{g})$-finite.
3. (If $\Gamma\backslash G$ is non-compact) they satisfy a certain condition of moderate growth at the cusps.

**Definition.** Functions on $\Gamma\backslash G$ with the above properties are called *automorphic forms* for $\Gamma\backslash G$.

Automorphic representations will be defined as irreducible subquotients of the space of automorphic forms, however in order to obtain a reasonable notion which doesn't include every representation, we have to give this definition later, when considering the action of a much larger group.

Automorphic forms do not necessarily belong to the space of an irreducible automorphic representation, or to a finite sum of such spaces. However, the following fundamental theorem of Harish-Chandra ensures, in particular, that the invariant subspace generated by an automorphic form is admissible and of finite length:

**Theorem 3.1** (Harish-Chandra)**.** *Fix a $K_\infty$-type (i.e. irreducible representation) $\tau$ and an ideal $\mathcal{J}$ of $\mathfrak{z}(\mathfrak{g})$ of finite codimension. Then $\mathcal{A}(\Gamma\backslash G)_{\mathcal{J},\tau}$, the $\tau$-isotypic component of the space of automorphic forms on $\Gamma\backslash G$ annihilated by $\mathcal{J}$, is finite-dimensional.*

## 28.4 Hecke operators

We return to the most "classical" setting, but from now on $\Gamma$ will be an arithmetic subgroup. Hecke observed that in this case the problem has more symmetries. For instance, let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Let $a \in \mathrm{GL}_2(\mathbb{Q})^+$, a rational matrix with

positive determinant. (Don't let yourselves be confused with the appearance of $\mathrm{GL}_2$; you can restrict you attention to $\mathrm{SL}_2(\mathbb{Q})$ if you want. But in fact, since the center acts trivially, we want to consider the elements of $\mathrm{PSL}_2(\mathbb{Q}) = \mathrm{PGL}_2(\mathbb{Q})$ which preserve the upper-half plane. These are precisely the images of elements in $\mathrm{GL}_2(\mathbb{Q})^+$, while $\mathrm{SL}_2(\mathbb{Q})$ does not account for all of them.) If $f$ is a function on $\Gamma\backslash\mathbb{H}$, or, equivalently, a $\Gamma$-invariant function on $\mathbb{H}$, then $L_a f$, defined as $L_a f(z) = f(az)$, is not $\Gamma$-invariant any more. However, it will be invariant by a subgroup which is commensurable to $\Gamma$. Now, consider the double coset $\Gamma a \Gamma$ and decompose it into (a finite number of) left cosets:

$$\Gamma a \Gamma = \sqcup_{i=1}^{n} \Gamma a_i.$$

We define the operation: $T_a : f \mapsto f|_a$ as: $f|_a = \sum_i = 1^n L_a f$. *Then $f|_a$ is $\Gamma$-invariant!* We call the subalgebra of $\mathrm{End}(\mathrm{Fns}(\Gamma\backslash\mathbb{H}))$ generated by the operators $T_a$ the *Hecke algebra*.

*Remark.* Later on, the name 'Hecke algebra' will be used for an abstract algebra which admits a natural homomorphism into $\mathrm{End}(\mathrm{Fns}(\Gamma\backslash\mathbb{H}))$, for every arithmetic subgroup $\Gamma$, whose image is the 'Hecke algebra' we just defined. But one should be aware of the fact that, in the literature, the term 'Hecke algebra' may refer to this subalgebra of $\mathrm{End}(\mathrm{Fns}(\Gamma\backslash\mathbb{H}))$.

**Fact.** *For $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, the algebra of Hecke operators is commutative.*

Now observe: The Hecke operators commute with the Laplacian! Hence, we can diagonalize them simultaneously and postulate:

> **3rd approximation:** Automorphic forms on $\Gamma\backslash\mathbb{H}$ are simultaneous eigenfunctions for $\Delta$ and the Hecke algebra on $\Gamma\backslash\mathbb{H}$.

We won't discuss here what happens when $\Gamma$ is not $\mathrm{SL}_2(\mathbb{Z})$, because this will be subsumed from our discussion of the adelic point of view.

## 28.5 Adelic formulation. Definition of automorphic representations.

Finally, we want to take into account the Hecke operators in a representation-theoretic way. For this, we need to move to the adelic picture. From now on, $G$ will not denote $\mathrm{SL}_2(\mathbb{R})$ but just $\mathrm{SL}_2$ as an algebraic group over $\mathbb{Q}$. As we saw in a previous lecture, if $K^\infty$ is a compact-open subgroup of the finite adeles $G(\mathbb{A}^\infty)$ of $G$ then $G(\mathbb{Q})\backslash G(\mathbb{A})/K^\infty$ is a finite union of spaces of the form $\Gamma\backslash G(\mathbb{R})$, as a $G(\mathbb{R})$-space. For simplicity, let us assume that there is only one such space, i.e.:

$$G(\mathbb{Q})\backslash G(\mathbb{A})/K^\infty = \Gamma\backslash G(\mathbb{R}).$$

We can do what we did in the real case: Lift our functions from $G(\mathbb{Q})\backslash G(\mathbb{A})/K^\infty$ to $G(\mathbb{Q})\backslash G(\mathbb{A})$, and consider the space that they generate under the $G(\mathbb{A})$-action. This is now a smooth representation of $G(\mathbb{A})$. What is the Hecke algebra in this

setting? Assume that $K^\infty = \prod_p K_p$ where $K_p = \mathrm{SL}_2(\mathbb{Z}_p)$. (This corresponds to the case of $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ which we discussed before.)

We saw in the previous lecture that $\mathcal{H}_p = \mathcal{H}(G(\mathbb{Q}_p), K_p)$ acts on $C^\infty(G(\mathbb{Q})\backslash G(\mathbb{A}))^{K_p}$. So, the algebra $\mathcal{H} := \otimes'_{p<\infty}\mathcal{H}_p$, where the restricted tensor product is taken with respect to the characteristic measures of the subgroups $K_p$, acts on $C^\infty(G(\mathbb{Q})\backslash G(\mathbb{A}))^{K_p}$. Its image in $\mathrm{End}(C^\infty(G(\mathbb{Q})\backslash G(\mathbb{A}))^{K_p})$ is the same as that of the operators $T_a$ defined previously (check!). The commutativity of the $T_a$'s follows from the Satake isomorphism which we will discuss below. The important part to keep in mind here is that, by our discussion of the Hecke algebra in the previous lecture, the information about the eigenvalues of the Hecke algebra is contained in the information about the representation of $G(\mathbb{A})$.

Now the space of automorphic forms (for all arithmetic subgroups $\Gamma$ together!) can be thought of as the space of functions on $G(k)\backslash G(\mathbb{A})$ which are:

1. Smooth under $G(\mathbb{A}^\infty)$ (i.e. invariant under a compact-open subgroup $K^\infty$),
2. $K_\infty$-finite,
3. $\mathfrak{z}(\mathfrak{g})$-finite and
4. (if $G$ is $k$-isotropic) satisfy a certain condition of moderate growth at the cusps (expressed in terms of coordinates of a Siegel domain).

**Definition.** An *automorphic representation* is an irreducible subquotient of the space of automorphic forms under the $G(\mathbb{A})$-action.

*Remark.* Strictly speaking, this definition makes no sense because there is no $G_\infty$ action but just $(\mathfrak{g}, K_\infty)$-action. But via the finiteness theorem of Harish-Chandra and the (Casselman-Wallach) theorem about equivalence between Harish-Chandra modules and admissible smooth representations of finite type, we can indeed talk about the corresponding smooth representation of $G(\mathbb{A})$.

In fact, it is more natural to think of automorphic representations as abstract (not necessarily irreducible) admissible, finite-type representations of $G(\mathbb{A})$ together with embeddings $\nu : \pi \hookrightarrow C(G(k)\backslash G(\mathbb{A}))$, because the images of those automatically satisfy the required conditions and generate the space of automorphic forms.

Finally, the question of "which representations are automorphic" is not obscurred by the issue of embedding vs. subquotient, because the question can be reduced to certain ("cuspidal") *embeddings*, as Langlands showed using his theory of Eisenstein series. But for now, this is probably too much technical discussion already, and we will come back to such issues when the time is ripe.

## 28.6 The unitary spectrum of $\mathrm{SL}_2(\mathbb{R})$; holomorphic modular forms.

# Chapter 29

# The Satake isomorphism and automorphic $L$-functions.

## 29.1 The tensor product theorem and unramified representations.

First we start with the tensor product theorem of Flath:

**Theorem 1.1.** *An irreducible admissible representation $\pi$ of $G(\mathbb{A})$ is uniquely isomorphic to $\otimes'_v \pi_v$ where:*

- *$\pi_v$ is an irreducible admissible representation of $G_v$.*
- *For almost all $v$, $\pi_v$ is unramified, and comes with a distinguished unramified vectore $u_v^0$.*
- *The restricted tensor product is taken with respect to the $u_v^0$'s.*

We need to explain the second condition. It only makes sense at non-archimedean places where $G$ is unramified. In those places, $G_v$ has one or more conjugacy classes of distinguished maximal compact subgroups $K_v$ called *special*. They satisfy the *Iwasawa decomposition* $G_v = P_v K_v$ where $P_v$ denotes (the $k_v$-points of) a minimal parabolic defined over $k_v$. For our purposes, it is enough to consider a smaller class of maximal compact subgroups called *hyperspecial*. By definition, a hyperspecial maximal compact subgroup of $G_v$ is a subgroup of the form $G(\mathfrak{o}_v)$, for some smooth model of $G$ over $\mathfrak{o}_v$. Such a model will not exist at every place, but will exist at almost every place if $G$ is defined over a global field.

More precisely, let $G$ be defined over a global field $k$. Then there exists a finite number of places $S$ such that $G$ has a smooth model over $\mathfrak{o}_S$ (the $S$-integers of $k$, i.e. integers outside of $S$) – this is obvious. Moreover, if we fix

two such models $G^1$ and $G^2$ and set $K_v^i = G^i(\mathfrak{o}_v)$ then $K_v^1 = K_v^2$ for almost every $v$. Therefore, for the (local) discussion that follows we will have fixed a "good" maximal compact open subgroup $K_v$ of $G_v$. Locally there may be many conjugacy classes of such subgroups, but we will remember that we started from a globally defined group $G$ and that almost all $K_v$ arose through the procedure we just described – in particular, the definitions which we will give will coincide at almost every place.

We drop the indices $v$ for now because the whole discussion will be over a non-archimedean local field (to be denoted simply by $k$). We are given an unramified reductive group $G$ over $k$ and a "good"(=special) maximal compact subgroup $K$. An irreducible smooth representation $\pi$ is called *unramified* if $\pi^K \neq 0$. This notion depends on the conjugacy class of $K$. By Theorem **??**, $\pi^K$ is an irreducible $\mathcal{H}(G,K)$-module. We will study the structure of the algebra $\mathcal{H}(G,K)$ to discover that it is commutative. It will follow that an irreducible $\mathcal{H}(G,K)$-module is one-dimensional, and hence to each irreducible unramified representation we can attach a character of the *spherical Hecke algebra* $\mathcal{H}(G,K)$.

## 29.2 The Satake isomorphism.

We continue in the same setting. For simplicity, let us assume first that $G$ is split. Let $A^*$ be the maximal torus of its dual group $\check{G}$. Hence, $\mathbb{C}[A^*]$ is the group algebra, over $\mathbb{C}$, of $\mathcal{X}(A)^*$, the co-character group of $A$. We have an action of the Weyl group $W$ on $A^*$ and hence on its coordinate ring. The *Satake isomorphism* states:

**Theorem 2.1.** *There is a canonical isomorphism:* $\mathcal{H}(G,K) \simeq \mathbb{C}[A^*]^W$.

The categorical quotient $A^*//W := \mathrm{spec}(\mathbb{C}[A^*]^W)$ is also equal to the geometric quotient $A^*/W$ (and is a non-singular variety), and therefore the characters of $\mathcal{H}(G,K)$ can be identified with points of $A^*/W$, or in other words:

**Corollary 2.2.** *The* spherical Hecke algebra $\mathcal{H}(G,K)$ *is commutative, and its characters are in canonical bijection with semisimple conjugacy classes in* $\check{G}$.

Before we discuss the general quasi-split case (which you may want to skip at first reading, anyway), let us discuss the split case a bit more. First of all, we notice that the dual group $\check{G}$ is essentially absent from the above formulation: Only its maximal torus and Weyl group are appearing. But different groups (for example, $\mathrm{Sp}_{2n}$ and $\mathrm{SO}_{2n+1}$) may have isomorphic maximal tori and Weyl groups.

The best way to put the dual group into the picture is by introducing its (algebraic) representations. First, recall the Chevalley isomorphism:

**Theorem 2.3** (Chevalley)**.** *If* $\check{G}$ *is a complex (connected) reductive group with maximal torus* $A^*$ *and Weyl group* $W$ *then restriction of regular functions to the maximal torus induces an isomorphism:* $\mathbb{C}[G]^G \simeq \mathbb{C}[A^*]^W$ *(where* $G$ *acts on the left side by conjugation).*

Now, a basis for conjugation-invariant rational functions on $G$ is given by characters of irreducible (highest-weight) algebraic representations. Those span a $\mathbb{Z}$-lattice inside of $\mathbb{C}[G]^G$ which can be identified with the algebra of isomorphism classes of *virtual* representations of $\check{G}$. (The algebra structure comes from tensor product, which is translated to multiplication of characters.)

There is a better version of the Satake isomorphism, where one starts with the $\mathbb{Z}$-lattice of integral-valued measures in $\mathcal{H}(G, K)$ and establishes a natural isomorphism with this $\mathbb{Z}$-lattice in $\mathbb{C}[G]^G$[1]. This makes clearer the relevance of $\check{G}$. For more details, see the article of Gross "On the Satake Isomorphism" in the Bulletin of the AMS.

A yet better isomorphism – because it is not of combinatorial nature – has been proven in the context of the Geometric Langlands program. (By Mircovic and Vilonen following, I understand, a suggestion of Drinfeld.) Here instead of algebras you have categories with extra (tensor, etc.) structure, of which the Hecke algebra – resp. the ring of invariant polynomials on $\check{G}$ – are just shadows. More precisely, the Hecke algebra is substituted by a category of perverse sheaves on the "affine Grassmannian" of $G$ (and the map back to the Hecke algebra is by Grothendieck's "function-sheaf correspondence"), and it is shown that this category has extra structure which makes it equivalent to the category of representations of a reductive group $\check{G}$. Hence, the group $\check{G}$ arises naturally from this category associated to $G$ and not vice-versa.[2]

**Unramified principal series.** While the above remarks were for general educational purposes and in order to read some foreign words, now we will discuss an aspect of the Satake isomorphism which is very basic to us: The torus $A^*$ can be identified with the torus of *unramified characters* of $A$ (where $A$ is a maximal, split torus of $G$). Here unramified means the same as above, but the maximal compact subgroup $A_0$ of $A$ can easily be described: It is the kernel of all characters of the form $a \mapsto |\chi(a)|$ where $\chi \in \mathcal{X}(A)$ and $|\bullet|$ denotes $p$-adic absolute value. Hence, unramified characters are of the form (non-uniquely): $|\chi_1|^{s_1} \cdot |\chi_2|^{s_2} \cdots |\chi_r|^{s_r}$, where the $\chi_i$'s are algebraic characters and $s_i \in \mathbb{C}$.

*Exercise.* Prove that $A^*$ can be canonically identified with the group of unramified characters of $A$.

Let $B$ be a Borel subgroup containing $A$. An unramified character $\chi$ of $A$ can be considered also as a character of $B$ via the quotient map: $B \twoheadrightarrow A$. The induced representation: $I(\chi) = \mathrm{Ind}_B^G(\chi\delta^{\frac{1}{2}}) = \{f : G \to \mathbb{C} | f(bg) = \chi\delta^{\frac{1}{2}}(b)f(g)\}$ (we will describe later what $\delta$ is, for now ignore it) is called an *unramified principal series*. Why is it unramified? Because of the Iwasawa decomposition: $G = BK$, there is a unique line of unramified vectors in $I(\chi)$, represented by functions which are constant on $K$.

[1] This is slightly imprecise, one needs a ring larger than $\mathbb{Z}$ to make it an isomorphism.

[2] The setting of the Geometric Langlands program does not include, unfortunately, a local or global field in characteristic zero, and hence the results are not easily transferable. Some very important bridges, however, have recently been created, which led, in particular, to the proof of the Fundamental Lemma.

We will discuss these representations in more detail later (their construction entails the following general feature: one understands representations of $G$ with the help of some basic "building blocks" + induction of the corresponding "building blocks" for Levi subgroups), for now I'm just writing some general properties: The representation $I(\chi)$ is not always irreducible, but it has a unique unramified irreducible subquotient which we will call $\pi_\chi$. (Why unique? Because "$K$-invariants" is an exact functor! If you want to start understanding these representations, try to understand th this statement and why it is true!) Moreover, $\pi_\chi$ is isomorphic to $\pi_{^w\chi}$ for every $w \in W$. So, let me hasten to tell you what is the explicit truth behind the Satake isomorphism: *The irreducible character of $\mathcal{H}(G,K)$ corresponding to a point $[\chi] \in A^*/W$ corresponds to the irreducible unramified representation $I(\chi)$ of $G$.*

We notice also the following straightforward reformulation: There is a bijection:

$$\left\{ \begin{array}{c} \text{semisimple conjugacy} \\ \text{classes in } \check{G} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{conjugacy classes of unramified} \\ \text{(i.e. inertia is in the kernel) homomorphisms} \\ \text{with semisimple image: } \mathcal{W}_k \to \check{G} \end{array} \right\}$$

where $\mathcal{W}_k$ is the Weil group of $k$ (not to be confused with $W_k$ which will stand for the relative Weyl group of $G$). The correspondence is simply:

$$(\text{conjugacy class of } g) \leftrightarrow (\text{conjugacy class of the map } \mathrm{Frob} \mapsto x).$$

In the general (quasi-split, but not necessarily split) case we let $P$ be a minimal $k$-parabolic of $G$ and $M$ a Levi subgroup of it. Hence, $M$ is the centralizer of a maximal $k$-split torus $S$. We let $W_k = \mathcal{N}(S)/\mathcal{Z}(S)$ be the relative Weyl group. It acts on the group of unramified characters of $M$. Let $X$ denote the group of these unramified characters; again, it naturally has the structure of a complex torus. Then:

**Theorem 2.4.** *There is a canonical (up to scalar multiplication) isomorphism: $\mathcal{H}(G,K) \simeq \mathbb{C}[X]^{W_k}$.*

Again, of course, the representation corresponding to a given point on $X/W_k$ is the representation induced from the corresponding character of $M$, considered as a character of $P$, i.e. the representation $I_P^G(\chi) := \mathrm{Ind}_P^G(\chi\delta_P^{\frac{1}{2}})$. (Again, $\delta_P$ will be explained below.) Moreover, there is a bijection:

$$X/W_k \leftrightarrow \left\{ \begin{array}{c} \text{conjugacy classes of unramified} \\ \text{(i.e. inertia is in the kernel) homomorphisms} \\ \text{with semisimple image: } \mathcal{W}_k \to {}^L G \text{ over } \mathcal{W}_k \end{array} \right\}$$

where "over $\mathcal{W}_k$" means that the composition $\mathcal{W}_k \to {}^L G \to \mathrm{Gal}(\bar{k}/k)$ is the identity on $\mathcal{W}_k$. (Recall that for unramified groups, $\mathrm{Gal}(\bar{k}/k)$ acts on $\check{G}$ via its unramified quotient.)

For more details, see the article of Cartier in the Corvallis proceedings.

## 29.3 Sketch of proof of the Satake isomorphism.

First, we explain the meaning of the modular character $\delta_P$: It is the absolute value of the modular character $\mathfrak{d}_P$ for $P = M \ltimes N$, look back at the section on Tamagawa measures. In other words, it is the quotient between right and left Haar measure on $P$. For $P = B$ we have denoted $\delta_B$ simply by $\delta$.

The idea of the proof is to let $\mathcal{H}(G, K)$ act faithfully on a space whose structure we understand. Having in the back of our heads that the eigenfunction of $\mathcal{H}(G, K)$ corresponding to a given $\chi \in X$ lives in $I_P^G(\chi)^K$, and thinking of $I_P^G(\chi)^K$ as a subspace of the space of functions on $N\backslash G$, the space in question will be related to $N\backslash G$.

More precisely, let $M_0$ denote the common kernel of all unramified characters of $M$ (equivalently, the maximal compact subgroup of $M$). The quotient $M/M_0$ is a lattice $\Lambda$, and the group ring $\mathbb{C}[\Lambda]$ is by definition the same as the coordinate ring $\mathbb{C}[X]$ (unfortunately the same notation for group rings and coordinate rings!). Let $\mathcal{M}(N\backslash G)$ denote the space of compactly supported, smooth measures on $N\backslash G$. The group $M \times G$ acts on it (with $M$ acting "on the left" and $G$ acting "on the right"). By convention, we normalize the action of $M$ so that it is "unitary": $m \cdot f(Nx) = \delta_P^{-\frac{1}{2}}(m) f(Nmx)$. (The word "unitary" makes sense only if we identify those measures with functions in $L^2$, by fixing a $G$-invariant measure on $N\backslash G$; in any case, by normalizing the action of $M$ this way on all possible spaces of functions, measures etc. on $N\backslash G$ we have the benefit that $C^\infty(N\backslash G)$ is the smooth dual of $\mathcal{M}(N\backslash G)$ under the $M \times G$ action.)

The subspace $\mathcal{M}(N\backslash G)^{M_0 \times K}$ is a module for $\mathcal{H}(M, M_0) \otimes \mathcal{H}(G, K)$. Clearly, $\mathcal{H}(M, M_0) \simeq \mathbb{C}[\Lambda]$ as an algebra. Moreover, by the Iwasawa decomposition: $N\backslash G/K = \Lambda$, therefore $\mathcal{M}(N\backslash G)^{M_0 \times K}$ can naturally be identified as a vector space with $\mathbb{C}[\Lambda]$. To make this identification $\mathcal{H}(M, M_0)$-equivariant, we normalize this identification as follows:

$$i : \mathcal{M}(N\backslash G)^{M_0 \times K} \ni \mu \mapsto i\mu(m) = \delta_P^{-\frac{1}{2}}(m)\mu(NmK) \in \mathcal{M}(N\backslash G)^{M_0 \times K}.$$

This way, the action map $\mathcal{H}(M, M_0) : h \mapsto h * 1_{N1K}$, where $1_{N1K}$ denotes the characteristic measure of $N1K$, gives rise to a commutative diagram:

$$\begin{array}{ccc} \mathcal{H}(M, M_0) & \longrightarrow & \mathbb{C}[\Lambda] \\ \downarrow & & \| \\ \mathcal{M}(N\backslash G)^{M_0 \times K} & \xrightarrow{i} & \mathbb{C}[\Lambda]. \end{array}$$

Up to now we only considered the action of $M$ to justify our statement that "we understand the structure of the space $\mathcal{M}(N\backslash G)^{M_0 \times K}$". Now let us consider the action of $\mathcal{H}(G, K)$, more precisely let $S : \mathcal{H}(G, K) \to \mathbb{C}[\Lambda]$ be the composition of the action map: $\mathcal{H}(G, K) \ni h \mapsto h * 1_{N1K}$ with the map $i$. ($S$ is called the *Satake map*.) Given that the actions of $\mathcal{H}(G, K)$ and $\mathcal{H}(M, M_0)$ commute, we get immediately:

$S$ is a homomorphism of rings.

Therefore, the Satake isomorphism rests upon showing that:

**Lemma 3.1.** *$S$ is injective and its image is $\mathbb{C}[\Lambda]^{W_k}$.*

Injectivity is easy; the statement about the image is less straightforward. See Cartier's article for more details.

## 29.4 Automorphic $L$-functions.

Let $k$ be a global field and $G$ a reductive group over $k$. Fix a finite set of primes $S$, which includes the archimedean primes and those primes where $G$ is ramified. By the tensor product theorem and the Satake isomorphism, for every collection $\{t_v\}_{v\notin S}$ of (twisted) semisimple conjugacy classes in $\check{G}$, we get a unique isomorphism class of everywhere unramified representations of $G(\mathbb{A}^S)$: $\pi = \otimes' \pi_{t_v}$, where $\pi_{t_v}$ is the unramified representation of $G_v$ with Satake parameter $t_v$. The question is: which such collections $\{t_v\}$ give rise to *automorphic* representations? It turns out that the requirement that a representation be automorphic imposes very strong conditions on the $t_v$'s. We certainly don't know or don't understand all those conditions, but we can express some of them in the language of $L$-functions.

Automorphic $L$-functions are defined as Euler products (i.e. products of the form $\prod_v L_v$, where $L_v$ is called the "local $L$-factor" and the product is over all places of $k$). The problem is that we don't know in all cases how to define the local $L$-factor, unless we are given the validity of the Local Langlands Conjecture. Therefore, we will define partial $L$-functions of the form: $L^S = \prod_{v\notin S} L_v$ and discuss properties of those. These won't cover all desired properties of automorphic $L$-functions (e.g. the generalized Riemann Hypothesis), but still the properties that they cover are already very important.

First, we discuss the case of the group $G = \mathrm{GL}_n$. Let $\pi = \otimes'_v \pi_v$ be an automorphic representation of $G$, and let $S$ be a finite set of places which includes the places where $G$ is ramified, the places where $\pi$ is ramified and the archimedean places. We define the *partial $L$-function* $L^S(\pi, s)$ as a function of one complex variable $s$ by the product: $L^S(\pi, s) = \prod_{v\notin S} L_v^S(\pi, s)$ for $\Re s \gg 0$, where $L_v$ depends only on $\pi_v$. More precisely, to $\pi_v$ corresponds by the Satake isomorphism a semisimple conjugacy class $t_v$ in $\mathrm{GL}_n(\mathbb{C})$, and if $\{a_1, a_2, \ldots, a_n\}$ are the eigenvalues of $t_v$ then we set:

$$L_v(s) = \prod_{i=1}^{n} \frac{1}{1 - a_n q^{-s}}.$$

Here is a very strong property that the $t_v$'s satisfy:

> $L^S(\pi, s)$ admits meromorphic continuation to the whole complex plane.

This was proven by Godement and Jacquet, whose proof was inspired by the thesis of Tate which treated the case $n = 1$. More is true; for instance the

$L$-function satisfies a functional equation. But this would require defining all local $L$-factors, and we will skip it for now. Some further expected properties of automorphic $L$-functions will be understood in the when we discuss the "Galois side" of the Langlands conjectures and Artin $L$-functions.

In fact, it is expected that one should be able to "play" with the coefficients $a_i$ and still produce $L$-functions with the same properties. This can be understood via the functoriality principle, which will be discussed in the next lecture, so for now we will confine ourselves to defining these *automorphic $L$-functions* which are expected to have (at least) meromorphic continuation.

Let $G$ be a reductive group over $k$, $\pi = \otimes'_v \pi_v$ an automorphic representation of $G$ and $\rho : {}^L G \to \mathrm{GL}_n(\mathbb{C}) \times \mathrm{Gal}(\bar{k}/k)$ an *algebraic* representation of its $L$-group over $\mathrm{Gal}(\bar{k}/k)$. We define the partial $L$-function $L^S(\pi, \rho, s) = \prod_{v \notin S} L_v(\pi, \rho, s)$ for $\Re s \gg 0$ where $L_v(\pi, \rho, s)$ depends only on $\pi_v$ and $\rho$ as follows: Let $T_v \in \mathrm{GL}_n(\mathbb{C})$ be the image of the Frobenius under the composite:

$$\mathrm{Frob} \mapsto {}^L G \to \mathrm{GL}_n$$

where the first map is the Satake map and the second is $\rho$. Then:

$$L_v(\pi, \rho, s) = \frac{1}{\det(I - q^{-s} T_v)}.$$

Of course, this definition coincides with the previous one when $G = \mathrm{GL}_n$ and $\rho$ is the identity representation.

Part of the Langlands conjectures is:

> $L(\pi, \rho, s)$ has meromorphic continuation to all $s$.

Concluding, it should be remarked that from the point of view of analytic number theory, $L$-functions are the most important objects and their coefficients such as $a_i$ (above) have arithmetic significance. The fact that combinations of those (expressed by the representations $\rho$ of the dual group) always give Euler products with analytic properties such as meromorphic continuation has very strong consequences, and I have heard Langlands say that "functoriality was invented to prove properties of $L$-functions".[3] Even proving a special case of "$L(\pi, \rho, s)$ has certain properties" (which in practice means proving a special case of functoriality) is considered to be a "big deal".

[3] At Arthur's birthday conference; possibly not an exact quote.

# Chapter 30

# The Langlands conjectures and arithmetic.

## 30.1 Weil groups and Weil-Deligne groups.

Let $k$ be a local field, and let $W_k$ denote its Weil group. Recall that class field theory establishes a canonical isomorphism:

$$W_k^{\mathrm{ab}} \simeq k^\times.$$

In fact, we have only explained what is the Weil group in the case of local non-archimedean fields. We give here the general definition: a *Weil group* for a local field $k$ is a topological group $W_k$ together with a continuous homomorphism: $W_k \to \mathrm{Gal}(\bar{k}/k)$ with continuous image, as well as an isomorphism $W_k^{\mathrm{ab}} \simeq k^\times$. It is required to satisfy certain natural conditions, for instance the composed map: $k^\times \to \mathrm{Gal}(\bar{k}/k)^{\mathrm{ab}}$ should be that of local class field theory. See the article of Tate in the Corvallis proceedings for details. A Weil group over a global field is defined in precisely the same way, except that one should substitute $k^\times$ with the idele class group $\mathbb{A}_k^\times/k^\times$. Weil groups exist and are unique up to unique isomorphism.

I remind that for local non-archimedean fields the Weil group can be considered as the preimage in $\mathrm{Gal}(\bar{k}/k)$ of $\langle \mathrm{Frob} \rangle$ under:

$$1 \to I_k \to \mathrm{Gal}(\bar{k}/k) \to \widehat{\langle \mathrm{Frob} \rangle} \to 1$$

where $I_k$ denotes inertia. The same essentially holds for global function fields: If $X$ is a smooth projective curve over $\mathbb{F}_q$ then one has a short exact sequence:

$$1 \to \mathrm{Gal}(\overline{\mathbb{F}_q(X)}/\overline{\mathbb{F}_q}(X))) \to \mathrm{Gal}(\overline{\mathbb{F}_q(X)}/\mathbb{F}_q(X)) \to \overline{\langle \mathrm{Frob} \rangle} \to 1$$

where Frob denotes the "geometric Frobenius" (which generates $\mathrm{Gal}(\overline{\mathbb{F}_q}(X)/\mathbb{F}_q(X))$). Again, the Weil group is the subgroup of the Galois group which maps to powers of the Frobenius.

If $k = \mathbb{C}$ then $W_k = \mathbb{C}^\times$. If $k = \mathbb{R}$ then $W_k$ is generated by $\mathbb{C}^\times$ and an element $j$ such that $j^2 = -1$ and $jzj^{-1} = \bar{z}$ for $z \in \mathbb{C}^\times$. For number fields there is no simple explicit way to describe the Weil group.

At first reading, one might find all these definitions too technical and unmotivated, and could think of the Weil group as some incarnation of the Galois group. The truth is that it contains more information than the Galois group: for instance in the case of a local non-archimedean field, we don't just have the Galois group of $\bar{k}^{\mathrm{ur}}/k$ (which is $\simeq \mathbb{Z}$) but also a distinguished topological generator for it, namely the Frobenius. And local class field theory is not just an isomorphism between the completions $\mathrm{Gal}(\bar{k}/k)^{\mathrm{ab}}$ and $\widehat{k^\times}$, but also between the distinguished dense subgroups $W_k$ and $k^\times$. Dually, if you want to parametrize representations (characters) of $k^\times$ you will use the Weil group, not its completion. The characters of $W_k$ which extend to the Galois group are precisely those of finite order, and the corresponding characters of $k^\times$ are called "of Galois type".

The truth is also that the Weil group is not the end of the story for global number fields. There should be some extension of it which parametrizes (automorphic) representations. But we will discuss this problem below. For now, I mention one more complication: In the case of non-archimedean local fields one needs to introduce the *Weil-Deligne* group $WD_k := W_k \ltimes \mathbb{G}_a(\mathbb{C})$ (I write $\mathbb{G}_a(\mathbb{C})$ instead of $\mathbb{C}$ to keep in mind that all representations we will encounter should be algebraic in this factor) where the semidirect product is defined by: $wzw^{-1} = |w|z$. Let $l$ be a prime different from the residue characteristic of $k$, and let $\check{G}$ be an algebraic group over $\bar{\mathbb{Q}}$. The following fact is true:

> There is a natural bijection between continuous representations: $W_k \to G(\bar{\mathbb{Q}}_l)$ and continuous representations (algebraic in the $G_a$-factor): $WD_k \to G(\mathbb{C})$.

Therefore, if you want, the natural setting might be that of $l$-adic representations, in which case there would be no need for the Weil-Deligne group. But since we prefer to work with complex representations, we introduce the Weil-Deligne group. (I also remark that while switching from complex to $l$-adic representations is very easy over a local, non-archimedean field, since we always take $l$ to be other than the residue characteristic, over a number field this is not the case and there is a very rich theory of $l$-adic representations which is vigorously being developed at the moment – usually called the *p-adic Langlands program*, where $p$ stands for $l$!.)

From now on, when we talk about "representation" of $W_k$ or $WD_k$ into some complex algebraic group, we will automatically and without further mentioning assume the following two conditions (except, of course, for continuity):

1. Any Frobenius element is mapped to a semisimple element. (Equivalently, all elements of $W_k$ have semisimple image.)

2. The representation is algebraic on the $\mathbb{G}_a$-factor of $WD_k$.

In what follows we will denote by $\mathcal{L}_k$ (for "Langlands group of the field") the Weil group of $k$, if $k$ is an archimedean local field, and the Weil-Deligne group of $k$, if $k$ is a non-archimedean local field.

## 30.2 Local Langlands Conjecture

The local Langlands conjecture generalizes the statement of local class field theory (in its dual form:

$$\{\text{Homomorphisms: } \mathcal{L}_k \to \mathbb{G}_m(\mathbb{C})\} \leftrightarrow \{\text{Irreducible representations of } \mathbb{G}_m(k)\})$$

to an arbitrary connected reductive group $G$, which (roughly speaking) takes the form of $\mathbb{G}_m$ on the right hand side, while its $L$-group ${}^LG$ takes the form of $\mathbb{G}_m$ on the left hand side.

Before we proceed, I will describe the picture for $\mathrm{PGL}_2$ over a non-archimedean local field. (Notice: discussing representations of $\mathrm{PGL}_2$ is exactly the same as discussing representations of $\mathrm{GL}_2$ with trivial central character.)

*Example* 2.1. Let $G = \mathrm{PGL}_2$ over $k$: a non-archimedean local field. We let $B$ denote a Borel subgroup, $A$ a maximal torus in $B$ (hence $A \simeq \mathbb{G}_m$) and $I(\chi) = \mathrm{Ind}_{B(k)}^{G(k)}(\chi\delta^{\frac{1}{2}})$, for any character $\chi$ of $A$ (thought of as a character of $B$). The representations $I(\chi)$ are irreducible with $I(\chi) \simeq I({}^w\chi)$, except when $\chi = \omega\delta^{\pm\frac{1}{2}}$, where $\omega$ is a quadratic character. In those cases, $I(\chi)$ has two subquotients, one of which is the one-dimensional space of the character $\omega$ (considered as a character of $\mathrm{PGL}_2(k)$), and the other is an infinite-dimensional representation which in the case of $\omega = 1$ is called the *Steinberg* or *special* representation. (For all $\omega$, let us call those the *Steinberg-type* representations.) Which of the two shows up as a subrepresentation, and which shows up as a quotient depends on whether we induce from $\omega\delta^{\frac{1}{2}}$ or from $\omega\delta^{-\frac{1}{2}}$.

To those representations we want to associate a homomorphism: $WD_k \to \mathrm{GL}_2(\mathbb{C})$ up to $\mathrm{GL}_2(\mathbb{C})$-conjugacy, called a *Langlands parameter*. First, to the unramified ones we will associate the homomorphism which is compatible with the Satake isomorphism: The group $WD_k$ has a cyclic quotient generated by Frobenius, and the map will be:

$$WD_k \to \langle \mathrm{Frob} \rangle \to \mathrm{GL}_2(\mathbb{C})$$

where Frob is mapped to an element in the conjugacy class of $\chi$. (Recall that unramified characters of $A(k)$ can be identified with elements of $A^*$, the maximal torus of the dual group.)

More generally, for every character $\chi$ of $A(k)$ we have by local class field theory a homomorphism: $W_k \to \mathbb{G}_m(\mathbb{C}) = A^*$ and the composite:

$$WD_k \to W_k \to A^*$$

is the map that we associate to $I(\chi)$, if it is irreducible, or to its one-dimensional subquotient, if it is reducible.

Finally, we need to describe the Langlands parameter for the Steinberg representation and the similar subquotients of $I(\omega\delta^{\frac{1}{2}})$. Recall that $WD_k = W_k \ltimes \mathbb{G}_a(\mathbb{C})$. We map again $W_k$ to $A^*$ according to local class field theory, and we map $\mathbb{G}_a$ to a unipotent subgroup normalized by $A^*$. Check that this is a homomorphism of groups!

*Remark.* In general, the additional "unipotent" parameter coming with the $\mathbb{G}_a$-factor of $WD_k$ has to do with the fact that a certain induced representation is reducible.

To finish the list of irreducible representations of $G(k)$, there are those which are not subquotients of any parabolically induced representation, called *supercuspidals*. We know now that those are in bijection with conjugacy classes of homomorphisms: $WD_k \to \mathrm{GL}_2(\mathbb{C})$, which factor through the quotient $W_k$ but such that *the image of* $W_k$ *is not contained in a Levi of* $\mathrm{GL}_2$ (with 'Levi', in this case, meaning just 'torus'). In other words, they are parametrized by something which is genuinely out of the scope of class field theory, and which sees deeper than the abelian quotient of $W_k$.

We return to a general $G$. To every irreducible admissible representation of $G(k)$ one should be able to associate a conjugacy class of representations

$$\phi : \mathcal{L}_k \to^L G$$

over $\mathrm{Gal}(\bar{k}/k)$ (i.e. compatible with the maps of both sides to $\mathrm{Gal}(\bar{k}/k)$), called a *Langlands parameter*. Recall that we have required from all representations of $\mathcal{L}_k$ to satisfy certain conditions, see the previous subsection.

If the group is not quasi-split, then the Langlands parameter should satisfy some extra conditions (recall that quasi-split forms and their inner forms have the same $L$-group). More precisely, the image of the Langlands parameters cannot lie inside the Levi of a proper (standard) parabolic of $\check{G}$, unless the dual of that parabolic is a $k$-rational parabolic of $G$. (Such Levi subgroups are called "relevant". The notion of "dual parabolic" has to do with the fact that standard parabolics are in bijection with subsets of the set of simple roots, and that the sets of simple roots of $G$ and $\check{G}$ are in bijection by passing from a root to its co-root.) For instance, for the multiplicative groups $D^\times$ of quaternion division algebras we only have Langlands parameters which are the analogs of supercuspidal and "Steinberg-type" parameters, not those corresponding to principal series. This is related to the fact that there are no $k$-rational Borel subgroups to induce from. Notice that even the trivial representation of $D^\times$ has ramified parameters; it turns out that its parameters are the same as those for the Steinberg representation of $\mathrm{GL}_2$!

In the case of $\mathrm{GL}_n$ this should be a one-to-one correspondence:

$$\{\text{Conjugacy classes of Langlands parameters } \phi : \mathcal{L}_k \to \mathrm{GL}_n(\mathbb{C})\}$$

$$\leftrightarrow \{\text{Isomorphism classes of irreducible admissible representations of } \mathrm{GL}_n(k)\}.$$

In the general case, this is no longer true. For every conjugacy class of Langlands parameters $\phi$ one should have a finite, non-empty set $\Pi_\phi$ of isomorphism

classes of irreducible representations (the *L-packet*), and these $L$-packets are a disjoint partition of the set of all irreducible representations. Representations in the same $L$-packet are called *L-indistinguishable* and the existence of $L$-packets is the dual phenomenon to the fact that the $k$-points of a $G(\bar{k})$-conjugacy class can split into several $G(k)$-conjugacy classes, although the relation between these two phenomena can only be understood when we talk about the trace formula.

Hence in the general case we write:

$$\{\text{Isomorphism classes of irreducible admissible representations of } G(k)\}$$

$$\to \{\text{Conjugacy classes of Langlands parameters } \phi : \mathcal{L}_k \to {}^L G\}.$$

It is essentially known how elements of an $L$-packet should be parametrized. See Vogan, *The Local Langlands Conjectures*.

## 30.3 The Global Langlands Conjecture

Let now $k$ be a global field, and $G$ a connected reductive group over $k$. There should be a "Langlands group" for $k$ which should be used to parametrize automorphic representations. For function fields, this is just the Weil group. For number fields we don't know what it is, but it is expected to be an extension of the Weil group. (Here I should say "extension of the Weil group by a compact group, but then I should have chosen a different – equivalent – incaranation of the Weil-Deligne group locally, and that would be confusing at this point.) It should be equipped with maps $\mathcal{L}_{k_v} \to \mathcal{L}_k$, for every place $v$, such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{L}_{k_v} & \longrightarrow & W_{k_v} \\ \downarrow & & \downarrow \\ \mathcal{L}_k & \longrightarrow & W_k. \end{array}$$

To each automorphic representation $\pi$ one should be able to attach a *Langlands parameter*:

$$\phi : \mathcal{L}_k \to {}^L G$$

over $\mathrm{Gal}(\bar{k}/k)$, satisfying again similar conditions as in the local case. It should have the property that if $\pi \simeq \otimes'_v \pi_v$ then the Langlands parameter for $\pi_v$ should be the composite: $\mathcal{L}_{k_v} \to \mathcal{L}_k \to {}^L G$. Vice versa, for every Langlands parameter one should have a non-empty packet of automorphic representations (the *L-packet*) and the set of all automorphic representations should be the disjoint union of all $L$-packets. Again, for $\mathrm{GL}_n$ the $L$-packets should be singletons.

This compatibility with the local Langlands correspondence is a very informative and meaningful statement, even when we don't completely know or understand the local Langlands correspondence. Indeed, recall that $\pi_v$ is unramified for almost every $v$, and for those representations we know the local

Langlands parameters. We notice that every local $L$-packet should have at most one unramified representation (for a given special maximal compact subgroup). Therefore, two automorphic representations in a global $L$-packet should have the same unramified local parameters almost everywhere.

## 30.4 Functoriality

Although the global Langlands conjecture cannot be properly formulated without knowing what the Langlands group should be, it already has some striking consequences which can be formulated quite precisely. Namely, let $G_1, G_2$ be two reductive groups and let $\rho$ be a homomorphism of their $L$-groups over $\mathrm{Gal}(\bar{k}/k)$:

$$\rho : {}^L G_1 \to {}^L G_2.$$

Then, composing every Langlands parameter into ${}^L G_1$ with $\rho$, we get a Langlands parameter into ${}^L G_2$. This tells us: *to every automorphic representation of $G_1$ should correspond, via $\rho$, an automorphic representation of $G_2$.* The word "correspond", here, has a very strong meaning by our knowledge of local Langlands for unramified representations. Indeed, by the Satake parameter the map $\rho$ provides an association $\pi_v \rightsquigarrow \Pi_v$ from unramified representations of $G_1(k_v)$ to unramified representations of $G_2(k_v)$ (for fixed special maximal compact subgroups). Therefore, if $\pi \simeq \otimes'_v \pi_v$ is an automorphic representation of $G_1$, then there should be an automorphic representation $\tau \simeq \otimes'_v \tau_v$ of $G_2$ with $\tau_v \simeq \Pi_v$ almost everywhere.

Notice that if $G_2 = \mathrm{GL}_n$ then we get an $n$-dimensional representation of ${}^L G_1$ over $\mathrm{Gal}(\bar{k}/k)$, and the $L$-function $L(\pi, \rho, s)$ of an automorphic representation $\pi$ of $G_1$ is equal to the "standard" $L$-function $L(\tau, s)$ of its functorial lift $\tau$ to $\mathrm{GL}_n$ (at least, the partial $L$-functions outside of a finite set of places $S$ are equal, since we have not discussed local $L$-factors for "bad" places yet). In particular, if functoriality holds then *every automorphic $L$-function is equal to a standard $L$-function for* $\mathrm{GL}_n$; in particular, the results of Godement and Jacquet hold. Hence, through functoriality one would establish good analytic properties for automorphic $L$-functions, e.g. meromorphic continuation.

Finally, we mention another "game" that one can play with Langlands parameters. The Langlands groups should behave like Galois groups, for instance if $E/k$ is an extension then $\mathcal{L}_E$ should be the preimage of $\mathrm{Gal}(\bar{k}/E) \subset \mathrm{Gal}(\bar{k}/k)$ in $\mathcal{L}_k$. Hence, if $\pi$ is an automorphic representation of $G$ over $k$, by restricting its Langlands parameter to $\mathcal{L}_E$ one gets a Langlads parameter for $G$ regarded as an algebraic group over $E$ (i.e. for $G \times_{\mathrm{spec}\, k} \mathrm{spec}\, E$), and hence to every automorphic representation of $G$ over $k$ should correspond (in the same sense as above) an automorphic representation of $G$ over $E$ called the *base change* of $\pi$. This is, in fact, a special case of functoriality (as I ask you to prove in the exercises).

# Chapter 31

# Complex representations of $p$-adic groups.

## 31.1 references

- W. Casselman, *Introduction to the theory of admissible representations of p-adic reductive groups.* Available online at: `http://www.math.ubc.ca/ cass/research/p-adic-book.dvi`.
- J. Bernstein, *Lectures on representations of reductive p-adic groups.* Notes by Karl Rumelhart. Available online at: `http://www.math.uchicago.edu/~mitya/langlands/Bernstein/Bernstein93new.dvi`.
- J. Bernstein, *Le "centre" de Bernstein.* Edited by P. Deligne. Travaux en Cours, Representations of reductive groups over a local field, 1–32, Hermann, Paris, 1984.
- P. Cartier, *Representations of p-adic groups: A survey.* In the Corvallis proceedings, Vol 1. Available online at: `http://www.ams.org/online_bks/pspum331/`.
- C. Moeglin, *Representations of* $\mathrm{GL}(n, F)$ *in the nonarchimedean case.* In: T. N. Bailey and A. W. Knapp, *Representation Theory and Automorphic forms*, Proceedings of Symposia in Pure and Applied Mathematics, AMS 1997.

Rami: The following outline is just an example – feel free to not follow it.

## 31.2 Introduction

Motivation from automorphic forms: We want to do harmonic analysis on the space $G(\mathbb{Q})\backslash G(\mathbb{A})$, i.e. to study functions on the space as a representation of

the group $G(\mathbb{A})$. Two steps: First understand abstract representations of $G(\mathbb{A})$, then try to understand which ones are *automorphic*, i.e. "show up" in functions on this space. Understanding representations of $G(\mathbb{A})$ is essentially the same as understanding representations of $G(k_v)$, for every completion $v$.

Historical remark: In the beginning, people were considering just the space $\Gamma\backslash G(\mathbb{R}) = G(\mathbb{Q})\backslash G(\mathbb{A})/K_f$ ($K_f$ a compact open subgroup of the finite adeles of $G$) as a $G(\mathbb{R})$-space: $K_\infty$-types, invariant differential operators (i.e. action of $U(\mathfrak{g})$). But they knew (for arithmetic $\Gamma$ only, not for arbitrary discrete subgroup!) that there are more symmetries to the problem, certain operators (*Hecke operators*) which commute with $G(\mathbb{R})$ action (for instance, commute with Laplacian=Casimir.) Gelfand and Graev first understood that there is an action of $G(\mathbb{Q}_p)$ hiding behind that.

What is the correct category of representations? The space $G(\mathbb{Q})\backslash G(\mathbb{A})$ carries an invariant measure, so one can study $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$; in particular we can restrict our attention to *unitary* representations. This turns out too difficult to do from the very beginning, will study a more "algebraic" category of (infinite-dimensional, in general) representations and then we can ask ourselves which ones are unitarizable.

## 31.3 Smooth representations

From now on, we think $G$ = the topological group of points of a reductive group over a local non-archimedean field.[1] Some general definitions:

Define smooth representation. Define Hecke algebra as the convolution algebra of locally constant, compactly supported distributions ($\Rightarrow$ measures) on $G$. Remark: Choice of Haar measure makes this isomorphic to smooth, compactly supported functions, but this is not a canonical way to think about them.

Hecke algebra is an *idempotented algebra*. Equivalence of categories between *smooth representations of* $G$ and *non-degenerate / unital modules for* $\mathcal{H}(G)$.

Remark (not for the class): The idempotents define projections, and you can take inverse limit $\hat{\mathcal{H}}$. These can be thought of as distributions on $G$ which, when convolved with an open compact subgroup, become compactly supported. The *center* of $G$ (the center of the category of smooth representations) can now canonically be identified with the center of $\hat{\mathcal{H}}$.

Schur's lemma. (Prove it if you want.) Remark: There is a bijection between the set of irreducible representations of $G$ and irreducible subspaces of $C^\infty(G)$ under the $G \times G$ action (via matrix coefficients), or (by dualizing) irreducible quotients of $C_c^\infty(G)$.

In fact, to dualize in the above statement, you need reflexivity of irreducible representations, so: Define admissible representations. Define contragradient (smooth dual) $\tilde{\pi}$ of $\pi$.

**Lemma 3.1.** *The following are equivalent:*

[1] But you can give the definitions in the generality you consider appropriate, for instance $l$-groups and $l$-spaces.

1. *$\pi$ is admissible.*

2. *$\tilde{\pi}$ is admissible.*

3. *$\tilde{\tilde{\pi}} = \pi$.*

Admissibility of irreducibles follows from their construction as subquotients of parabolically induced. Example of non-admissible: The space of smooth, compactly supported functions on a non-compact $G$-space.

## 31.4 Induction, restriction.

Define induced, compactly induced representations. Induction is right-adjoint to restriction (Frobenius reciprocity), but in general there is no left-adjoint. Compact induction is left-adjoint if subgroup is compact open. For parabolic subgroups, induction and compact induction agree.

Parabolic induction (twist by $\delta_P^{\frac{1}{2}}$ to preserve unitarity) and Jacquet restriction. Preserves admissibility. Supercuspidals $\Leftrightarrow$ matrix coefficients have compact support modulo center $\Leftrightarrow$ both injective and projective in the category of smooth $\omega$-representations, where $\omega$ is the central character.

Talk about the Bernstein center.

## 31.5 Intertwining operators

Talk about distributions on $l$-spaces and show an example of how we find intertwining operators between parabolically induced representations: Write the filtration of your space by orbits, show how to compute the Jacquet module of the graded piece corresponding to each orbit, show how you can represent the equivariant distribution on that orbit by an integral and say that a "natural" way to extend this distribution to the whole space is by making sense of the same integral: It will typically converge somewhere, and have rational continuation. Do not do that at the level of general statements, just an example, e.g. principal series on $\mathrm{GL}_n$.

# Appendix A

# Some category theory

## A.1 Some category theory

This is by no means any complete account of any of the topics – consult a textbook such as Gelfand and Manin, *Methods of homological algebra*. We view category theory as just a language which allows us to avoid repeating the same arguments over and over again – hence the abstract definitions are less important than the actual properties.

## A.2 Some universal objects

We start by reminding the universal properties defining some objects. (As usual, a universal property defines the object, if it exists, uniquely up to unique isomorphism.)

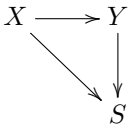
### A.2.1

The *(direct) product* $X \times Y$ is an object $Z$ together with morphisms $Z \to X$, $Z \to Y$ such that "any pair of morphisms $Z' \to X, Z' \to Y$ factors uniquely throught $Z$". (The expression in the quotation marks means that there is a unique morphism $Z' \to Z$ such that the given morphisms are obtained via the obvious compositions.)

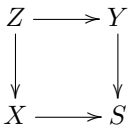
### A.2.2

The *fiber product* $X \times_S Y$ of a diagram $X \to S \leftarrow Y$ ("fiber product of $X$ and $Y$ over $S$") "is" the direct product in a new category $\mathcal{C}_S$ whose objects are

morphisms $X \to S$ and whose morphisms are commutative diagrams:



We put the word "is" in quotation marks because we think of the fiber product as an object $Z$ in $\mathcal{C}$ with morphisms $X \leftarrow Z \to Y$ such that the compositions with the maps to $S$ coincide. A diagram:



where $Z$ is the fiber product of $X$ and $Y$ is called *cartesian*. (So, notice that "cartesian" is much stronger than "commutative"; it's not enough that the composition of the arrows is the same over all paths.)

## A.2.3

The *coproduct* $X \cup Y$ "is" the product in the opposite category $\mathcal{C}^o$. (Similarly: the *fibered coproduct* or *pushout* $X \cup_S Y$, the notion of *cocartesian square* etc.)

*Exercise.* Describe those objects in the categories of sets, vector spaces, topological spaces, and Banach spaces.

## A.2.4

An *epimorphism* is a morphism $f : A \to B$ with the property that for any commutative diagram:

$$X \xrightarrow{f} Y \underset{g_2}{\overset{g_1}{\rightrightarrows}} Z$$

we have $g_1 = g_2$.

A *monomorphism* is a morphism which is an epimorphism in the opposite category.

An isomorphism is both a monomorphism and an epimorphism, but the opposite is not always true. Example: in the category of topological spaces, the map:

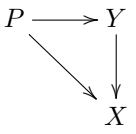
$$[0, 1) \ni \theta \mapsto e^{2\pi i\theta} \in S^1.$$

## A.2.5

An object $P$ is *projective* if the functor:

$$\mathrm{Hom}(P, \bullet) : \mathcal{C} \to Sets$$

preserves epimorphisms, that is: every epimorphism $Y \to X$ lifts to a morphism: $P \to Y$. ("Lifts" means that the resulting diagram:




is commutative.)

An object is *injective* if it is projective in the opposite category.

### A.2.6

To be added: limits and colimits.

## A.3 Abelian categories

### A.3.1 Axiom 1

The sets $\mathrm{Hom}(X, Y)$ are abelian groups, and the composition of morphisms is bi-additive (i.e. satisfies the distributive law on both the left and the right).

### A.3.2 Axiom 2

There exists a zero object 0 in $\mathcal{C}$, that is, an object such that $\mathrm{Hom}(0, 0)$ is the zero group. (This implies that both $\mathrm{Hom}(0, X)$ and $\mathrm{Hom}(X, 0)$ are zero for all objects $X$.)

### A.3.3

Assume that Axioms 1 and 2 are satisfied. The *kernel* of a morphism $X \xrightarrow{\phi} Y$ is a morphism $K \xrightarrow{k} X$ such that $\phi \otimes k = 0$ and which is a final object for the category of these morphisms. (That is, for any $K' \xrightarrow{k'} X$ there is a unique $K' \to K$ with $k' = k \circ h$.)

A *cokernel* is a kernel in the opposite category.

*Exercise.* Define kernels and cokernels in the categories of vector spaces and topological abelian groups (or Banach spaces, if you prefer).

### A.3.4 Axiom 3

For any two objects $X_1, X_2$ there exist an object $Y$ and morphisms $X_1 \underset{i_1}{\overset{p_1}{\leftrightarrows}} Y \underset{i_2}{\overset{p_2}{\leftrightarrows}} X_2$ such that $p_1 i_1 = \mathrm{Id}_{X_1}, p_2 i_2 = \mathrm{Id}_{X_2}, i_1 p_1 + i_2 p_2 = \mathrm{Id}_Y$, $p_2 i_1 = p_1 i_2 = 0$.

A corollary of the axiom is: $Y$ is both the product and the coproduct of $X_1$ and $X_2$. However, the axiom is slightly stronger than that.

### A.3.5

A category which satisfies Axioms 1–3 is called an *additive category.*

### A.3.6 Axiom 4

Kernels and cokernels exist, and moreover any morphism $\phi : X \to Y$ fits into a sequence: $K \xrightarrow{k} X \xrightarrow{i} I \xrightarrow{j} Y \xrightarrow{k'} K'$ with the properties:

1. $j \circ i = \phi$;
2. $K \to X$ is the kernel of $\phi$, and $Y \to K'$ is the cokernel;
3. $I$ is both the kernel of $j$ and the cokernel of $i$.

*Remark.* Think of $I$ as the "image" of $\phi$.

### A.3.7 Monomorphisms and epimorphisms

An additive category which satisfies Axiom 4 is called *abelian.* We will now see some properties of abelian categories – from now on all categories are abelian without mentioning it:

**Lemma 3.1.** *A morphism is a monomorphism iff its kernel is zero, and an epimorphism iff its cokernel is zero. If both the kernel and cokernel are zero, then it is an isomorphism.*

*Exercise.* Show that this fails for the category of Banach spaces (or, more simply, the category of topological abelian groups). Which of the axioms fails in this case, and how?

### A.3.8 Simple objects

A simple object in an abelian category is an object A such that any monomorphism $B \to A$ is either 0 or an isomorphism.

**Lemma 3.2.** *If $A$ is a simple object then any morphism $m : B \to A$ is either 0 or an epimorphism.*

*Proof.* If we split $m$ as in Axiom 4, we have:

$$0 \to K \to B \to I \to A \to K' \to 0.$$

Then $I \to A$ is a monomorphism, hence either an isomorphism (in which case $K'$ is zero and $m$ is an epimorphism) or zero (in which case $m$ is zero). □

*Exercise.* Show that we could have defined simple objects in the dual way, i.e. applying this definition to the opposite category.

**Corollary 3.3** (of the lemma and the exercise)**.** *Let $A, B$ be simple objects. Then either they are isomorphic, or any morphism: $A \to B$ is zero.*

### A.3.9 The splitting lemma

A sequence $\cdots \to X_i \xrightarrow{\phi_i} X_{i+1} \xrightarrow{\phi_{i+1}} X_{i+2} \to \cdots$ is called *exact* if, for all $i$, $\ker(\phi_{i+1}) = \operatorname{coker}(\phi_i)$.

**Lemma 3.4.** *For short exact sequence* $0 \to A \xrightarrow{\alpha} B \xrightarrow{\beta} C \to 0$ *the following are equivalent:*

- $\alpha$ *splits (i.e. there is a morphism* $a : B \to A$ *such that* $a \circ \alpha = \mathrm{Id}$*);*
- $\beta$ *splits (i.e. there is a morphism* $b : C \to B$ *such that* $\beta \circ b = \mathrm{Id}$*);*
- $B$ *is isomorphic to* $A \cup C$ *(direct product, or direct sum, is the same since the category is abelian).*

*Proof.* Let us first prove that the first implies the second: Given a splitting of $\alpha$, we first construct a morphism: $b' : B \to B$ as follows: $b' = \mathrm{Id}_B - \alpha \circ a$. (Recall that the Hom sets are abelian groups.)

The composition of $\alpha$ and $b'$ is zero. By the universal property of cokernels (since $C$ is the cokernel of the map $B \to A$), we have a unique $b : C \to B$ such that $b' = b \circ \beta$.

We now show that $\beta \circ b = \mathrm{Id}_C$. The morphisms $\beta = \mathrm{Id}_C \circ \beta$ and $\beta \circ b \circ \beta$:

$$B \to C \rightrightarrows C$$

coincide since:

$$\beta \circ b \circ \beta = \beta \circ b' = \beta \circ (\mathrm{Id}_B - \alpha \circ a) = \beta \circ \mathrm{Id}_B,$$

given that $\beta \circ \alpha = 0$.

But since $B \to C$ is an epimorphism, this implies that the two arrows from $C$ to $C$ coincide, that is: $\beta \circ b = \mathrm{Id}_C$.

The fact that the second statement implies the first follows in exactly the same manner.

Now we prove the third part assuming the other two (the converse is easy):

From $a$ and $\beta$ we get a morphism: $B \xrightarrow{j} A \cup C$ (by the universal property of direct products). We claim that it is a monomorphism. Indeed, assume that $D \xrightarrow{\phi} B$ has the property that its composition with $j$ is zero. That means that $\beta \circ \phi$ is zero; but then $\phi$ factors through $\alpha$, i.e. $\phi = \alpha \circ \phi'$ for some $\phi' : D \to A$ (since $A$ is the kernel of $\beta$); but we also have $a \circ \phi = a \circ \alpha \circ \phi' = \phi'$ is equal to zero, hence $\phi$ is zero. This implies that the kernel of $B \to A \cup C$ is zero.

Similarly, we can prove that $j$ is an epimorphism (exercise!). Therefore it is an isomorphism! □

### A.3.10 Jordan–Hölder theorem

A *finite composition series* for an object $A$ in an abelian category is a sequence of monomorphisms:

$$0 = A_0 \to A_1 \to A_2 \to \cdots A_n = A,$$

such that the *graded pieces* $(\operatorname{gr} A)_i := A_i/A_{i-1}$ are all simple objects (hence non-zero). (For simplicity of notation, we have denoted here by $A_i/A_{i-1}$ the cokernel of $A_{i-1} \to A_i$.)

An object which admits a finite composition series is said to be *of finite length*. Its *length* is a well-defined integer, as the following theorem shows.

**Theorem 3.5** (Jordan–Hölder)**.** *If $A$ is an object of finite length, then all composition series of $A$ have the same length, and the (unordered) set of simple objects*

$$\mathrm{JH}(A) := \{(\operatorname{gr} A)_i\}_{i=1}^{n}$$

*(the* Jordan–Hölder content *of $A$) is "the same" (=bijection of isomorphic objects) for all composition series of $A$.*

*Proof.* Exercise! □