HIGHER ALGEBRA 1 & 2 (MATH 570 & 571)
COURSE NOTES
FALL 2022 & WINTER 2023
VERSION: May 15, 2025

EYAL Z. GOREN,
MCGILL UNIVERSITY
EYAL.GOREN@MCGILL.CA

Contents

Part 1. GENERAL CONSTRUCTIONS AND TECHNIQUES

1. Categories and functors

1.1. **The concept of a category.** A **category C** consists of the following data:

- A collection of **objects** $ob(\mathbf{C})$.
- For any two objects A, B in $ob(\mathbf{C})$, a set of **morphisms**
$$\mathrm{Mor}(A, B).$$
- For any three objects A, B, C of $ob(\mathbf{C})$, a **composition** map
$$\mathrm{Mor}(B, C) \times \mathrm{Mor}(A, B) \to \mathrm{Mor}(A, C), \qquad (g, f) \mapsto g \circ f.$$
- For any object A a morphism $1_A \in \mathrm{Mor}(A, A)$.

And subject to the following:

- For all $f \in \mathrm{Mor}(A, B), g \in \mathrm{Mor}(B, C), h \in \mathrm{Mor}(C, D)$
$$h \circ (g \circ f) = (h \circ g) \circ f.$$
- For all $f \in \mathrm{Mor}(A, B)$
$$f \circ 1_A = f, \quad 1_B \circ f = f.$$

Remark 1.1.1. We note that

(1) 1_A is unique. If $1'_A$ is another morphism with the same properties as 1_A then
$$1_A = 1_A \circ 1'_A = 1'_A.$$
(2) $f \in \mathrm{Mor}(A, B)$ is called an **isomorphism** if $\exists g \in \mathrm{Mor}(B, A)$ such that $g \circ f = 1_A$ and $f \circ g = 1_B$. Such g, if it exists, is unique: if g' has the same properties then
$$g = g \circ 1_B = g \circ (f \circ g') = (g \circ f) \circ g' = 1_A \circ g' = g'.$$

An object A of $\mathbf{C}$ is called **initial** (respectively, **final**) if for any object B, including the object A itself, the set $\mathrm{Mor}(A, B)$ (respectively, $\mathrm{Mor}(B, A)$) has a single element. An initial (respectively, final) object, if it exists, is unique up to unique isomorphism. Namely, if A is an initial object and also A' is an initial object then the unique morphism $f \in \mathrm{Mor}(A, A')$ is an isomorphism. And similarly for final objects. We leave that as an exercise. An object is called a **zero** object if it is both initial and final.

1.1.1. *Examples.* Here are examples of categories that will be very important to us.

(1) **Sets**. The objects are sets and morphisms are functions. 1_A, for a set A, is the identity function from A to A. Any set with one element is a final object. The empty set is an initial object.
(2) $_{\mathbf{R}}\mathbf{Mod}$ and $\mathbf{Mod}_{\mathbf{R}}$. Let R be a ring (always associative with 1, but not necessarily commutative). The category $_{\mathbf{R}}\mathbf{Mod}$ (respectively, $\mathbf{Mod}_{\mathbf{R}}$) is the category of **left** R-modules (resp., **right** R-modules) M. Namely, M is an abelian group and there is a function
$$R \times M \to M, \quad (r, m) \mapsto rm$$

such that (1) $(r_1+r_2)m = r_1m + r_2m$, (2) $(r_1r_2)m = r_1(r_2m)$, (3) $1\,m = m$ and (4) $r(m_1+m_2) = rm_1 + rm_2$, where $r, r_1, r_2 \in R$, $m, m_1, m_2 \in M$.

The definition of a right R-module is entirely analogous, but now we write mr and have axioms of the sort $(mr_1)r_2 = m(r_1r_2)$.

A **morphism** between two left R-modules M, N, is a function

$$f\colon M \to N,$$

which is a homomorphism of groups and also satisfies $f(rm) = rf(m)$, for all $r \in R, m \in M$. (For right R-modules we require $f(mr) = f(m)r$.)

(3) **Gps**. The objects are groups and the morphisms are group homomorphisms.

(4) **Top**. The objects are topological spaces $(X, \mathcal{T}_X)$ and morphisms are continuous maps. Recall that a **topological space** is a set X together with a collection $\mathcal{T}_X$ of subsets of X (called **open** sets) that is closed under arbitrary unions and finite intersections, and such that $\emptyset$ and X are open sets. The collection $\mathcal{T}_X$ is called a **topology** on X.

A function $f\colon X \to Y$ between topological spaces is called **continuous** if for every $U \in \mathcal{T}_Y$, $f^{-1}(U) := \{x \in X : f(x) \in U\}$ belongs to $\mathcal{T}_X$.

(5) **Posets**. A **poset** A is a partially order set, namely a set for which for some elements x, y we have that $x \leq y$. We require that always $x \leq x$ and that if $x \leq y$ and $y \leq z$ then $x \leq z$. We also require that if $x \leq y$ and $y \leq x$ then $x = y$.

For example, if we take the positive integers and we define that $x \leq y$ if $x|y$, then we get a poset. If we take the set of all non-zero integers and define $x \leq y$ if $x|y$, we do not get a poset (what fails?). Note that there is no requirement that for any x, y either $x \leq y$ or $y \leq x$ and, indeed, in our example the numbers 3 and 5 cannot be compared.

A morphism $f\colon X \to Y$ of posets is a function f such that whenever $x_1, x_2 \in X$ and $x_1 \leq x_2$ then $f(x_1) \leq f(x_2)$.

In a category $\mathbf{C}$, a morphism $f\colon A \to B$ is called a **monomorphism** if for every object X and morphisms $g_1, g_2\colon X \to A$, $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$. It is called an **epimorphism** if for every object X and morphisms $g_1, g_2\colon B \to X$, $g_1 \circ f = g_2 \circ f$ implies $g_1 = g_2$. For example, in the category of sets a morphism is mono if and only if it is injective, and it is epi if and only if it is surjective. However, the natural inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ in the category of rings is both mono and epi, but is not an isomorphism.

1.2. **The concept of a functor.** Let $\mathbf{C}, \mathbf{D}$ be categories. A **covariant functor** (respectively, **conravariant functor**)

$$F\colon \mathbf{C} \to \mathbf{D}$$

is a rule associating to every object A of $\mathbf{C}$ an object FA of $\mathbf{D}$, and to any morphism $f \in \mathrm{Mor}_{\mathbf{C}}(A, B)$ a morphism $Ff \in \mathrm{Mor}_{\mathbf{D}}(FA, FB)$ (respectively, $Ff \in \mathrm{Mor}_{\mathbf{D}}(FB, FA)$), such that the following hold:

- $F1_A = 1_{FA}$;
- $F(g \circ f) = Fg \circ Ff$ (respectively, $F(g \circ f) = Ff \circ Fg$).

A functor F is called **faithful** if for any $A, B \in ob(\mathbf{C})$ and $f_1, f_2 \in \mathrm{Mor}_{\mathbf{C}}(A, B)$ we have

$$Ff_1 = Ff_2 \Longrightarrow f_1 = f_2;$$

namely, F is injective on morphisms. F is called **full** if it is surjective on morphisms:

$$F\colon \mathrm{Mor}_{\mathbf{C}}(A, B) \twoheadrightarrow \mathrm{Mor}_{\mathbf{D}}(FA, FB)$$

(respectively, $F\colon \mathrm{Mor}_{\mathbf{C}}(A,B) \twoheadrightarrow \mathrm{Mor}_{\mathbf{D}}(FB,FA)$).

1.2.1. *Examples.*

(1) **Forgetful functors.** These are functors that forget some of the information in the source category. A simple instance is the functor

$$\Phi\colon \mathbf{Gps} \to \mathbf{Sets},$$

where $\Phi G = G$, now viewed only as a set, and $\Phi f = f$, now viewed merely as a function. This is a covariant, faithful, but not full, functor. There are many other kind of forgetful functors. For example, there is a natural functor ${}_{\mathbf{R}}\mathbf{Mod} \to \mathbf{AbGps}$ associating to an R-module its underlying abelian group and viewing any module homomorphism merely as a group homomorphism, and so on.

(2) **Abelianization**. The definition of the category $\mathbf{AbGps}$ is quite clear (objects = abelian groups, morphisms = group homomorphisms). We have a functor

$$F\colon \mathbf{Gps} \to \mathbf{AbGps},$$

that takes a group G to its abelianization

$$FG = G^{ab} = G/G',$$

where G' is the commutator subgroup. Let us denote an element of G^{ab} by $\bar{g}$. Then to a homomorphism $f\colon G \to H$ we associate a homomorphism

$$Ff\colon G^{ab} \to H^{ab}, \quad Ff(\bar{g}) = \overline{f(g)}.$$

We leave as an exercise to show that Ff is a well-defined homomorphism and that F is a functor that is covariant, but neither full nor faithful.

(3) If G is a group let $*_G$ be the category with one object, denoted $*$, and

$$\mathrm{Mor}(*,*) := G$$

where composition is simply multiplication in the group $g \circ f := gf$. If H is a another group, a covariant functor $F\colon *_G \to *_H$ is the same thing as a homomorphism $G \to H$.

(4) ${}_{\mathbf{k}}\mathbf{VSp}$. Let k be a field. Then the category ${}_{\mathbf{k}}\mathbf{Mod}$ is what we normally call the category of k-vector spaces and we denote it ${}_{\mathbf{k}}\mathbf{VSp}$; we also denote $\mathrm{Mor}(V,W)$ by $\mathrm{Hom}_k(V,W)$. In particular, morphisms are just k-linear maps. Given a vector space V, the **dual** vector space V^* is defined as

$$\mathrm{Hom}_k(V,k),$$

the space of k-linear functionals (k-linear maps from V to k). If $T\colon V \to W$ is a linear map, we have the dual map $T^*\colon W^* \to V^*$ defined by

$$(T^*\phi)(v) = \phi(Tv), \quad \phi \in W^*, v \in V.$$

The association $V \mapsto V^*, T \mapsto T^*$ defines a contravariant functor ${}_{\mathbf{k}}\mathbf{VSp} \to {}_{\mathbf{k}}\mathbf{VSp}$.

(5) The group ring $k[G]$. Let G be a group and k a field. The **group ring** $k[G]$ can be described as

$$k[G] = \left\{ \sum_{g\in G} a_g[g] : a_g \in k, g \in G, \text{ all but finitely many } a_g \text{ are zero} \right\}.$$

Here the sums are formal sums and $[g]$ is a formal symbol associated to an element $g \in G$. So we can also think about $k[G]$ as the direct sum of copies of k, indexed by elements of G. Or, from yet

another perspective, as functions $G \to k$ that are non-zero on finitely many elements. We define

$$(\sum_{g\in G} a_g[g]) + (\sum_{g\in G} b_g[g]) = \sum_{g\in G}(a_g + b_g)[g],$$

and

$$(\sum_{g\in G} a_g[g])(\sum_{g\in G} b_g[g]) = \sum_{g\in G}(\sum_{h\in G} a_h b_{h^{-1}g})[g].$$

This is a (non-commutative, in general) ring. In the interpretation of $k[G]$ as functions, we can think about the product as convolution of functions. Further, $G \to k[G]$ is functor from the category of groups to the category of rings.

We shall sometimes use the notation e_g for $[g]$; mainly for "psychological reasons". A general element of $\mathbb{C}[G]$ will then be written as $\sum_{g\in G} a_g e_g$.

(6) $\mathbf{Rep}(\mathbf{G})$. Let G be a finite group. A **complex representation** of G is a pair (ρ, V) where V is a finite dimensional complex vector space and $\rho\colon G \to \mathrm{GL}(V)$ is a homomorphism of groups. The complex representations are the objects of the category $\mathbf{Rep}(\mathbf{G})$. The morphisms are

$$\mathrm{Hom}_G((\rho, V), (\tau, W)) := \{T\colon V \to W : T \circ \rho(g) = \tau(g) \circ T, \forall g \in G\}.$$

For example, $\mathbb{C}[G]$ is a complex representation of G, the **regular representation**, where

$$\rho(g)(\sum_{h\in G} a_h[h]) = \sum_{h\in G} a_h[gh] = [g](\sum_{h\in G} a_h[h]).$$

(7) Any category $\mathbf{C}$ has the **identity functor** $\mathbb{1}_{\mathbf{C}}$ that takes any object, and any morphism, to itself. It is of course covariant, full and faithful.

(8) $\mathbf{Hom}$. Let R be a ring and fix some $A \in ob({}_{\mathbf{R}}\mathbf{Mod})$. Define functors:

$$F = \mathrm{Hom}_R(A, -)\colon {}_{\mathbf{R}}\mathbf{Mod} \to \mathbf{AbGps} \qquad \text{(covariant)}$$
$$G = \mathrm{Hom}_R(-, A)\colon {}_{\mathbf{R}}\mathbf{Mod} \to \mathbf{AbGps} \quad \text{(contravariant)}$$

Namely, for any $B \in {}_{\mathbf{R}}\mathbf{Mod}$, $\mathrm{Hom}_R(A, B)$ (resp., $\mathrm{Hom}_R(B, A)$) is an abelian group and any $f\colon B \to C$ induces a group homomorphism

$$Ff\colon \mathrm{Hom}_R(A, B) \to \mathrm{Hom}_R(A, C), \quad Ff(g) = f \circ g;$$
$$Gf\colon \mathrm{Hom}_R(C, A) \to \mathrm{Hom}_R(B, A), \quad Ff(g) = g \circ f.$$

(9) A composition of two functors is a functor. If the two functors have the same variance then the composition is covariant.

1.3. **Morphisms of functors and equivalence of categories.** Let $F, G\colon \mathbf{C} \to \mathbf{D}$ be functors of the same variance, i.e., either both covariant, or both contravariant. A **morphism of functors** (or, a **natural transformation**)

$$\varphi\colon F \to G,$$

is a collection of morphisms $\{\varphi_A\}$,

$$\varphi_A \in \mathrm{Mor}_{\mathbf{D}}(FA, GA), \quad \forall A \in ob(\mathbf{C}),$$

such that for all $f \in \mathrm{Mor}_{\mathbf{C}}(A,B)$, the following diagram commutes in the covariant case:

$$\begin{array}{ccc} FA & \xrightarrow{\varphi_A} & GA \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle Gf} \\ FB & \xrightarrow[\varphi_B]{} & GB \end{array}$$

In the contravariant case, we require instead the commutativity of the following diagram (the vertical arrow now point up)

$$\begin{array}{ccc} FA & \xrightarrow{\varphi_A} & GA \\ {\scriptstyle Ff}\uparrow & & \uparrow{\scriptstyle Gf} \\ FB & \xrightarrow[\varphi_B]{} & GB \end{array}$$

We say that F is **isomorphic** to G (or that F is **naturally equivalent** to G) if there is such a morphism φ such that all φ_A are isomorphisms. We then write $F \cong G$.

Two categories $\mathbf{C}, \mathbf{D}$ are called (anti)-equivalent if there exists a covariant (respectively, contravariant) functors

$$F\colon \mathbf{C} \to \mathbf{D}, \quad \mathbf{G}\colon \mathbf{D} \to \mathbf{C},$$

such that

$$G \circ F \cong \mathbb{1}_{\mathbf{C}}, \quad F \circ G \cong \mathbb{1}_{\mathbf{D}}.$$

For example, $G \circ F \cong \mathbb{1}_{\mathbf{C}}$ means that for every $A \in ob(\mathbf{C})$ and for any $f \in \mathrm{Mor}_{\mathbf{C}}(A,B)$ we have a commutative diagram:

$$\begin{array}{ccc} GFA & \xrightarrow[\sim]{\varphi_A} & A \\ {\scriptstyle GFf}\downarrow & & \downarrow{\scriptstyle f} \\ GFB & \xrightarrow[\sim]{\varphi_B} & B. \end{array}$$

Typically, GFA is not actually *equal* to A, but we are given a natural way to *"identify"* them with the morphism φ_A.

Example 1.3.1. Suppose that $\mathbf{C} = \mathbf{D}$ is the category of finite dimensional k-vector spaces. Then the duality functor F from $\mathbf{C}$ to itself, $V \mapsto V^*, T \mapsto T^*$, is an anti-equivalence. That is, for every finite-dimensional k-vector space V there is a natural isomorphism

$$\varphi_V : V \to FFV = V^{**}, \quad v \mapsto \{\phi \mapsto \phi(v)\},$$

and under this the transformation T^{**} is identified with T.

A functor $F\colon \mathbf{C} \to \mathbf{D}$ is called **essentially surjective** if for any object B of $\mathbf{D}$ there is an object A of $\mathbf{C}$ and an isomorphism $FA \cong B$. The following theorem is extremely useful in proving equivalence of categories. The proof is not very enlightening, and is not more complicated in the general case than our Example 1.3.4 below. We omit it.

Theorem 1.3.2. *A functor $F\colon \mathbf{C} \to \mathbf{D}$ is an (anti) equivalence of categories, i.e., there is a functor $G\colon \mathbf{D} \to \mathbf{C}$, such that $G \circ F \cong \mathbb{1}_{\mathbf{C}}, F \circ G \cong \mathbb{1}_{\mathbf{D}}$, if and only if F is fully-faithful and essentially surjective.*

Example 1.3.3. Let G be a finite group. The category $\mathbf{Rep}(\mathbf{G})$ of finite-dimensional complex representations of G is equivalent to the category $\mathbf{fg_{\mathbb{C}[G]}Mod}$ of finitely-generated left $\mathbb{C}[G]$-modules:

Let (ρ, V) be a finite-dimensional representation of G. We make V into a $\mathbb{C}[G]$-module by defining for $v \in V$,

$$(\sum_{g\in G} a_g[g]) * v = \sum_{g\in G} a_g\ \rho(g)(v).$$

Note that if $\{v_i\}$ generate V as a $\mathbb{C}$-vector space, they generate V as a $\mathbb{C}[G]$-module, as we have $\mathbb{C} \hookrightarrow \mathbb{C}[G]$ by $z \mapsto z[1]$. A map $T\colon (\rho, V) \to (\tau, W)$ of representations becomes a map of $\mathbb{C}[G]$-modules as

$$T((\sum_{g\in G} a_g[g]) * v) = T(\sum_{g\in G} a_g\ \rho(g)(v)) = \sum_{g\in G} a_g\ T(\rho(g)(v))$$
$$= \sum_{g\in G} a_g\ \tau(g)(T(v)) = (\sum_{g\in G} a_g[g]) * T(v).$$

Conversely, given a $\mathbb{C}[G]$-module M, where we denote multiplication of a module element m by a ring element as $(\sum_{g\in G} a_g[g]) * m$, we make M into a $\mathbb{C}$-vector space by defining $zm = z[1] * m$, namely, by using the natural inclusion of $\mathbb{C}$ in $\mathbb{C}[G]$. If $\{m_i : i = 1, \dots, n\}$ generate M as a $\mathbb{C}[G]$-module then $\{[g] * m_i : g \in G, i = 1, \dots, n\}$ generate M as a $\mathbb{C}$-vector space. Maps of $\mathbb{C}[G]$-modules induce maps of representations, by a similar calculation.

In this example, the composition of the two functors, in either order, is the identity functor.

Example 1.3.4. Let k be a field and consider a category $\mathbf{C}$ whose objects are k^n for $n = 0, 1, 2, \dots$ and where morphisms are

$$\mathrm{Hom}(k^b, k^a) = M_{a,b}(k)$$

(matrices of size $a \times b$ with entries in k). Composition of morphisms is simply matrix multiplication.

On the other hand, we have the category $\mathbf{fg_k VSp}$ of finite dimensional k-vector spaces with morphisms being linear maps, and there is a natural functor

$$F\colon \mathbf{C} \to \mathbf{fg_k VSp},$$

where $F(k^n) = k^n$ and where we view a matrix in $M \in M_{a,b}(k)$ as a k-linear map $k^b \to k^a$, $v \mapsto Mv$. One of the first results we prove in linear algebra is that this functor is essentially surjective (namely, every vector space has a basis, etc.) and fully-faithful which means that we can identify $\mathrm{Hom}_k(k^b, k^a)$ with $M_{a,b}(k)$.

Although this is a very simple example, it shows an important feature: *an equivalence of categories can be very far from being a bijection on the level of objects.* The objects in $\mathbf{C}$ are a set of cardinality $\aleph_0$, while the collection of objects of $\mathbf{fg_k VSp}$ is not even a set.[1]

A much deeper example of equivalence of categories is provided by Morita's equivalence.

1.4. **Morita equivalence.** Let R be a ring, not necessarily commutative. Then the set of $n \times n$ matrices with entries from R, $M_n(R)$, is also a ring under the usual matrix operation, only that one has to be very methodic keeping right right and left left. Namely,

$$(a_{ij})(b_{ij}) = (c_{ij}),$$

where

$$c_{ij} = \sum_{\ell} a_{i\ell} b_{\ell j},$$

[1]We do not explain it further; here, as well as throughout the text, we will attempt to disregard any foundational difficulties unless absolutely necessary.

which might be different than $\sum_\ell b_{\ell j}a_{i\ell}$. We view R as contained in $M_n(R)$ via

$$R \to M_n(R), \quad r \mapsto r \cdot I_n.$$

Denote by E_{ij} the $n \times n$ matrix all whose entries are zero, except the ij entry which is 1. These matrices have two important properties:

- for $r \in R$ we have $rE_{ij} = E_{ij}r$;
- Let δ_{ki} is Kronecker's delta function, then

$$E_{\ell k}E_{ij} = \delta_{ki}E_{\ell j}.$$

As an application of our criterion for equivalence of categories, we prove the following theorem, which is part of what's called **Morita equivalence**.

Theorem 1.4.1 (Morita)**.** *Let R be a ring and let $n \geq 1$ an integer. The categories $_{\mathbf{R}}\mathbf{Mod}$ and $_{\mathbf{M_n(R)}}\mathbf{Mod}$ are equivalent.*

Proof. Define

$$F\colon {}_{\mathbf{R}}\mathbf{Mod} \to {}_{\mathbf{M_n(R)}}\mathbf{Mod}$$

by

$$F(M) = M^n$$

on objects M, and

$$F(f) = {}^t(f, f, \dots, f),$$

for a morphism $f\colon M \to N$. Here we think about M^n as columns vectors with entries in M and denote an element $(m_1, \dots, m_n)$ of it by $\underline{m}$. The action of $M_n(R)$ is given by the usual formula

$$(a_{ij})_{i,j=1}^n \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} = \begin{pmatrix} \sum_\ell a_{1\ell}m_\ell \\ \vdots \\ \sum_\ell a_{n\ell}m_\ell \end{pmatrix}.$$

It is obvious that F is a faithful functor.

We prove that F is a full functor. Any morphism $\varphi\colon M^n \to N^n$ of $M_n(R)$-modules is of the form

$$\varphi(\underline{m}) = {}^t(\varphi_1(\underline{m}), \dots, \varphi_n(\underline{m})),$$

and the φ_i are maps of R-modules. We therefore have

$$\varphi({}^t(m, 0, \dots, 0)) = \varphi(E_{11}{}^t(m, 0, \dots, 0)) = E_{11}\varphi({}^t(m, 0, \dots, 0)) = {}^t(\varphi_1(m, 0, \dots, 0), 0, \dots, 0).$$

This implies that φ_i vanishes on ${}^t(m, 0, \dots, 0)$ for all $i \neq 1$. Similarly, we conclude that φ_i vanishes on ${}^t(0, \dots, 0, \underset{j}{m}, 0, \dots, 0)$ for any $i \neq j$. Using additivity of φ, it follows that

$$\varphi({}^t(m_1, \dots, m_n)) = (\varphi_1(m_1), \dots, \varphi_n(m_n)).$$

(Where by $\varphi_j(m_j)$ we mean $\varphi_j({}^t(0, \dots, 0, \underset{j}{m_j}, 0, \dots, 0))$.)

Let $\sigma \in S_n$ and $E(\sigma)$ the matrix $\rho^{St}(\sigma^{-1})$ so that

$$E(\sigma) \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} = \begin{pmatrix} m_{\sigma(1)} \\ \vdots \\ m_{\sigma(n)} \end{pmatrix}.$$

Then, on the one hand,

$$\varphi(E(\sigma)^t(m_1,\dots,m_n)) = {}^t(\varphi_1(m_{\sigma(1)}),\dots,\varphi_n(m_{\sigma(n)})),$$

while on the other hand,

$$\varphi(E(\sigma)^t(m_1,\dots,m_n)) = E(\sigma)\varphi({}^t(m_1,\dots,m_n)) = (\varphi_{\sigma(1)}(m_{\sigma(1)}),\dots,\varphi_{\sigma(n)}(m_{\sigma(n)}).$$

As this holds for all σ and every $m_i \in M$, we conclude that

$$\varphi_1 = \cdots = \varphi_n.$$

That is,

$$\varphi = F(\varphi_1),$$

and F is a full functor.

It remains to prove that F is essentially surjective and this is the most subtle part. Let M be an $M_n(R)$-module. We need to show that $M \cong N^n$ as $M_n(R)$-modules, for some R-module N. The proof will show that $N = E_{11}M$ works.

Claim 1. We have an isomorphism of left R-modules

$$M \cong E_{11}\cdot M \oplus \cdots \oplus E_{nn}\cdot M.$$

To prove that consider the functions

$$f\colon M \to E_{11}\cdot M \oplus \cdots \oplus E_{nn}\cdot M, \quad m \mapsto (E_{11}m,\dots,E_{nn}m)$$

and

$$g\colon E_{11}\cdot M \oplus \cdots \oplus E_{nn}\cdot M \to M, \quad (a_1,\dots,a_n) \mapsto a_1+\cdots+a_n.$$

We claim that these maps, that are homomorphisms of R-modules, are mutual inverses. We have

$$g(f(m)) = \sum_i E_{ii}\cdot m = (\sum_i E_{ii})\cdot m = 1\cdot m = m.$$

Writing $a_i = E_{ii}a_i'$ we find that

$$\begin{aligned} f(g(a_1,\dots,a_n)) &= (E_{11}\cdot(\sum_j E_{jj}a_j'),\dots,E_{nn}\cdot(\sum_j E_{jj}a_j')) \\ &= (\sum_j E_{11}E_{jj}a_j',\dots,\sum_j E_{nn}E_{jj}a_j') \\ &= (E_{11}a_1',\dots,E_{nn}a_n') \\ &= (a_1,\dots,a_n). \end{aligned}$$

Claim 2. We have equalities $E_{\ell 1}M = E_{\ell\ell}M$.

Indeed, as $E_{\ell 1} = E_{\ell\ell}E_{\ell 1}$, we have $E_{\ell 1}M = E_{\ell\ell}(E_{\ell 1}M) \subseteq E_{\ell\ell}M$. On the other hand, as $E_{\ell\ell} = E_{\ell 1}E_{1\ell}$, we have $E_{\ell\ell}M = E_{\ell 1}(E_{1\ell}M) \subseteq E_{\ell 1}M$.

We thus conclude that

$$E_{11}M \oplus \cdots \oplus E_{n1}M \cong M, \quad (a_1,\dots,a_n) \mapsto \sum a_i.$$

Claim 3. We have isomorphisms $E_{11}M \cong E_{\ell 1}M$ by the map $a \overset{f}{\mapsto} E_{\ell 1}a$.

The map $a \mapsto E_{\ell 1}a$ is restriction of the R-module homomorphism $M \to E_{\ell 1}M$ given by the same formula. Therefore, it is an R-module homomorphism. We claim that the map

$$E_{\ell 1}M \to E_{11}M, \quad b \overset{g}{\mapsto} E_{1\ell}b,$$

is a well-defined inverse map. Let us write $a = E_{11}a' \in E_{11}M, b = E_{\ell 1}b' \in E_{\ell 1}M$. We first note that $g(b) = E_{1\ell}E_{\ell 1}b' = E_{11}b' \in E_{11}M$ and so g is well-defined. Now,

$$g(f(a)) = g(f(E_{11}a')) = E_{1\ell}E_{\ell 1}E_{11}a' = E_{11}a' = a.$$

Also,

$$f(g(b)) = E_{\ell 1}E_{1\ell}E_{\ell 1}b' = E_{\ell 1}b' = b.$$

Collecting the three claims together, we conclude that we have an isomorphism

$$\varphi : F(E_{11}M) = E_{11}M \oplus \cdots \oplus E_{11}M \to M,$$

given by the map

$$(a_1, \ldots, a_n) \mapsto (E_{11}a_1, \ldots, E_{n1}a_n) \mapsto \sum_{\ell=1}^{n} E_{\ell 1}a_\ell.$$

But this is so far an isomorphism of R-modules. We have to check that this isomorphism is an isomorphism of $M_n(R)$-modules. As $M_n(R)$ is spanned over R by the matrices E_{ij} and $(E_{11}M)^n$ is spanned over R by vectors of the form ${}^t(0, \ldots, 0, a, 0, \ldots, 0)$, it is enough to check that

$$\varphi(E_{\ell k}{}^t(0, \ldots, 0, a, 0, \ldots, 0)) = E_{\ell k}\varphi({}^t(0, \ldots, 0, a, 0, \ldots, 0)),$$

where a is at the i-th coordinate.

If $i \neq k$ then $E_{\ell k}{}^t(0, \ldots, 0, a, 0, \ldots, 0) = 0$ and so $\varphi(E_{\ell k}{}^t(0, \ldots, 0, a, 0, \ldots, 0)) = 0$. Also, $\varphi({}^t(0, \ldots, 0, a, 0, \ldots, 0)) = E_{i1}a$ and so $E_{\ell k}\varphi({}^t(0, \ldots, 0, a, 0, \ldots, 0)) = E_{\ell k}E_{i1}a = 0$.

If $i = k$ then $\varphi(E_{\ell i}{}^t(0, \ldots, 0, \underset{i}{a}, 0, \ldots, 0)) = \varphi({}^t(0, \ldots, 0, \underset{\ell}{a}, 0, \ldots, 0)) = E_{\ell 1}a$, and on the other hand, $E_{\ell i}\varphi({}^t(0, \ldots, 0, a, 0, \ldots, 0)) = E_{\ell i}E_{i1}a = E_{\ell 1}a$. □

Example 1.4.2. Morita equivalence can be used to provide a quick and conceptual proof of the following statement. Let R be a ring, n, k positive integers. Consider $M_{n,k}(R)$ as a left-$M_n(R)$ module. Then its endomorphisms as an $M_n(R)$-module are $M_k(R)$, acting from the right.

$$\mathrm{End}_{M_n(R)}(M_{n,k}(R)) = M_k(R).$$

1.4.1. *Division algebras.* A classical application of Morita's equivalence is when one can completely classify the R-modules. You can amuse yourself by considering finitely generated $\mathbb{Z}$-modules. We will discuss an extreme case when R is a division ring. Recall that a **division ring**, or a **skew field**, is a ring R in which $0 \neq 1$ and every non-zero element is a **unit** of R. That is, if $x \neq 0$ is an element of R then there exists $y \in R$ such that $xy = yx = 1$. The simplest example is a field, of course.

If R is a division ring then the centre of R is a field and R is an algebra over it. It thus makes sense to already discuss division algebras R over a field K, meaning that R is a division ring containing K in its centre. The problem of classifying division algebras over a field K is a complicated and difficult one. Assume that R is of finite dimension over K. Here are some interesting results:

- If K is a finite field then R is a finite field as well. See Theorem 1.4.3.

- If K is the real numbers, R is either $\mathbb{R}, \mathbb{C}$ or the **Hamilton quaternions**

$$\mathbb{H} = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}k$$

with $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$. In particular, there are no division algebras whose dimension over $\mathbb{C}$ is 9, say.

- Suppose that the characteristic of K is not 2. There is then a general construction of 4-dimensional K-algebras called **quaternion algebras**.[2] Let $a, b \in K^\times$ and define an algebra R, often denoted $\left(\frac{a,b}{k}\right)$, by

$$K \oplus Ki \oplus Kj \oplus Kk,$$

where $i^2 = a, j^2 = b, ij = -ji = k$. Define the **norm** of an element $u = x + yi + zj + wk$, $x, y, z, w \in K$, of R by

$$N(u) = N(x + yi + zj + wk) = x^2 - ay^2 - bz^2 + abw^2.$$

We have the following properties of the norm function:
(1) $N(u) = u\bar{u}$, where $\bar{u} = x - yi - zj - wk$;
(2) $N(u_1u_2) = N(u_1)N(u_2)$.

Using this, it is not hard to prove that u is invertible if and only if $N(u) \neq 0$. Thus, R is a division algebra if and only if the quadratic form $x^2 - ay^2 - bz^2 + abw^2$ does not represent 0 (this means that the only solution to $x^2 - ay^2 - bz^2 + abw^2 = 0$ is $x = y = z = w = 0$). For example, if $K = \mathbb{R}$ and $a = b = -1$ we get the form $x^2 + y^2 + z^2 + w^2$ and conclude that the Hamilton quaternions $\mathbb{H}$ are indeed a division algebra.

- Over $\mathbb{Q}$ there are "plenty" of non-isomorphic non-commutative division algebras. For example, for every positive integer n there is a division algebra whose centre is precisely $\mathbb{Q}$ and whose dimension over $\mathbb{Q}$ is n^2. A more general notion is that of a central simple algebra over $\mathbb{Q}$, or a field K. It turns out that there is a group whose elements are central simple algebras R over K, up to an equivalence $R \sim M_n(R)$ for every n. It is called the **Brauer group**. In this group every quaternion algebra over K has order equal to 2, unless it is isomorphic to $M_2(K)$, in which case it has order 1. The group structure is given by the tensor product over K. See § 29.4.

If R is a division algebra, the theory of R-modules is rather similar to the theory of vector spaces initially. Things get complicated once we get to determinants, eigenvectors, eigenvalues, and so-on due to the non-commutativity of R, but initially there is hardly a difference. For example, the definition of linearly independent sets, spanning sets and so on, are the same. There are the same characterizations of a basis and Steinitz substitution lemma works and provides the invariance of dimension. It requires nothing but patience to prove that every finitely generated R-module is isomorphic to R^d for some integer $d \geq 0$ that is uniquely determined. Morita equivalence now gives that every finitely generated $M_n(R)$-module is of the form $(R^d)^n \cong M_{n,d}(R)$ for some d. Otherwise said, every $M_n(R)$-module is isomorphic to a direct sum of R^n, where R^n is viewed as column vectors of length n with entries in R.

1.4.2. *Wedderburn's little theorem.*

Theorem 1.4.3 (Wedderburn)**.** *Let R be a finite division ring then R is a field.*

[2] Quaternion algebras can also be defined over a field of characteristic 2, but the construction is a bit different.

Proof. (E. Witt) Let R be a finite division ring. We prove the theorem by induction on the cardinality of R; the case $|R| = 2$ is clear.

Let K be the centre of R. It is a field with q elements. Let $n = \dim_K(R)$. Our goal is to prove that $n = 1$. Suppose that $n > 1$. Let $r \in R \setminus K$ and let K_r be its centralizer in R. It is easy to check that K_r is a division algebra as well and that we have inclusions

$$K \subseteq K_r \subsetneqq R.$$

Thus, using induction, K_r is a field too. If we let $n_r = \dim_K(K_r)$ we have $n_r | n$ and $1 \leq n_r < n$. The multiplicative groups $K^\times, K_r^\times, R^\times$, have orders $q-1, q^{n_r}-1$, and $q^n - 1$, respectively.

Write the class equation for the multiplicative group $R^\times$ acting on itself by conjugation:

$$q^n - 1 = q - 1 + \sum_{r \notin K} \frac{q^n - 1}{q^{n_r} - 1}, \tag{1}$$

where the summation is over representatives for the conjugacy classes of elements $r \notin K$.

The idea now is to find an integer dividing $\frac{q^n-1}{q^{n_r}-1}$ for all r and therefore $q-1$. We will show that this integer is larger than $q-1$, thereby arriving at a contradiction.

For a positive integer b, let $\Phi_b(x) = \prod_{\zeta \text{ prim. root of order } b}(x - \zeta) \in \mathbb{Z}[x]$ denote the (monic) cyclotomic polynomial of degree $\varphi(b)$ (where φ is Euler's function). Then

$$x^n - 1 = \prod_{d|n} \Phi_d(x), \qquad x^{n_r} - 1 = \prod_{d|n_r} \Phi_d(x).$$

As $n_r | n$ we have $\Phi_n(x) | \frac{x^n-1}{x^{n_r}-1}$. Substituting q for x, we find that

$$\Phi_n(q) | \frac{q^n - 1}{q^{n_r} - 1}.$$

By the class equation (1), the integer $\Phi_n(q)$ divides $q-1$ as well. On the other hand

$$\Phi_n(q) = \prod_{\zeta} (q - \zeta),$$

where the product is taken over all $\varphi(n)$ roots of unity of order n. As the absolute value of each such complex number $q - \zeta$ is greater than $q - 1$ we find $|\Phi_n(q)| > q - 1$. Contradiction. □

1.5. **Adjoint functors.** Let $F\colon \mathbf{C} \to \mathbf{D}$ and $G\colon \mathbf{D} \to \mathbf{C}$ be functors of the same variance. We say that (F, G) is an **adjoint pair** (and that F is **left-adjoint** to G, and G **right-adjoint** to F) if the following holds. For every $A \in ob(\mathbf{C}), B \in ob(\mathbf{D})$ we are given isomorphisms (of sets)

$$\mathrm{Mor}_{\mathbf{D}}(FA, B) \xrightarrow[\sim]{\varphi_{A,B}} \mathrm{Mor}_{\mathbf{C}}(A, GB)$$

such that for all $f \in \mathrm{Mor}_{\mathbf{C}}(A, A'), g \in \mathrm{Mor}_{\mathbf{D}}(B, B')$ the following diagrams commute, in the covariant case (in the contravariant case, replace the direction of the vertical arrows):

$$\begin{array}{ccc} \mathrm{Mor}_{\mathbf{D}}(FA, B) & \xrightarrow{\varphi_{A,B}} & \mathrm{Mor}_{\mathbf{C}}(A, GB) \\ {\scriptstyle (-)\circ Ff}\uparrow & & \uparrow{\scriptstyle (-)\circ f} \\ \mathrm{Mor}_{\mathbf{D}}(FA', B) & \xrightarrow{\varphi_{A',B}} & \mathrm{Mor}_{\mathbf{C}}(A', GB) \end{array} \qquad \begin{array}{ccc} \mathrm{Mor}_{\mathbf{D}}(FA, B) & \xrightarrow{\varphi_{A,B}} & \mathrm{Mor}_{\mathbf{C}}(A, GB) \\ {\scriptstyle g\circ(-)}\downarrow & & \downarrow{\scriptstyle Gg\circ(-)} \\ \mathrm{Mor}_{\mathbf{D}}(FA, B') & \xrightarrow{\varphi_{A,B'}} & \mathrm{Mor}_{\mathbf{C}}(A, GB') \end{array}$$

We say that the isomorphisms $\varphi_{A,B}$ are **natural**.

Remark 1.5.1. We will introduce later the notions of direct and inverse limits. It is a theorem that if (F, G) is an adjoint pair of functors then F preserves all direct limits and G preserves all inverse limits. Many algebraic constructions can be viewed as limits, so this is a very desirable property. For more on that, see § 4.3.

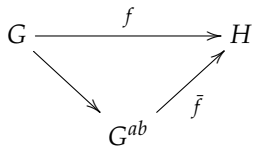
Example 1.5.2. Let $inc\colon \mathbf{AbGps} \to \mathbf{Gps}$ be the inclusion functor of the category of abelian groups in the category of groups, and let $ab\colon \mathbf{Gps} \to \mathbf{AbGps}$ be the abelianization functor as in §1.2.1. Namely,

$$inc(G) = G,\ inc(f) = f; \qquad ab(G) = G^{ab},\ ab(f)(\bar{g}) = \overline{f(g)}.$$

Then (ab, inc) is an adjoint pair of covariant functors: For a group G and an abelian group H we have

$$\mathrm{Hom}_{\mathbf{AbGps}}(G^{ab}, H) \cong \mathrm{Hom}_{\mathbf{Gps}}(G, H),$$

because every homomorphism $f\colon G \to H$ factors uniquely as



Of course, one needs to check that this identification is natural relative to group homomorphisms $G \to G', H \to H'$. This is not hard and is left as an exercise.

Example 1.5.3. Free construction. A left-adjoint to a forgetful functor is a **free construction** functor. This is more of a principle than a definition, and we make the statement precise in one example.

For any set S, let $F(S)$ be the **free group** on the alphabet S. Recall that elements of $F(S)$ are **words** $\omega_1^{\epsilon_1} \cdots \omega_n^{\epsilon_n}$ where $\omega_i \in S$ and $\epsilon_i \in \{\pm 1\}$; two words are identified if we can get from one to the other by canceling, or adding, expressions of the form $\omega^{\epsilon}\omega^{-\epsilon}$ where $\omega \in S$ and $\epsilon \in \{\pm 1\}$. Multiplication is defined as concatenation. For example, if $S = \{x, y\}$, the elements of $F(S)$ include $x, y, x^{-1}yxyy^{-1}x$, etc. Note that the last word is the same as $x^{-1}yxx$ and $yy^{-1}x^{-1}yxx$. We usually use the short hand $x^{-1}yx^2$ for this word. We usually write 1 to denote all the words equivalent to the empty word; this element is the identity element of $F(S)$.

Any function $f\colon S \to T$ induces a group homomorphism

$$F(f)\colon F(S) \to F(T), \quad F(f)(\omega_1^{\epsilon_1} \dots \omega_n^{\epsilon_n}) = f(\omega_1)^{\epsilon_1} \cdots f(\omega_n)^{\epsilon_n}.$$

In this way we get a functor

$$F\colon \mathbf{Sets} \to \mathbf{Gps}.$$

We let Φ be the forgetful functor $\mathbf{Gps} \to \mathbf{Sets}$. Then (F, Φ) is an adjoint pair:

$$\mathrm{Hom}_{\mathbf{Gps}}(F(S), H) \xrightarrow{\sim} \mathrm{Hom}_{\mathbf{Sets}}(S, H)$$

2. Tensor products

2.1. **Basic definitions and goals.** Let R be a ring, and let $A \in \mathbf{Mod_R}$ and $B \in {}_{\mathbf{R}}\mathbf{Mod}$ be modules.[3] An R-**biadditive** map is a function

$$f\colon A \times B \to H,$$

where H is an abelian group, such that for all $a, a_1, a_2 \in A,\ b, b_1, b_2 \in B,\ r \in R$:

(1) $f(a_1 + a_2, b) = f(a_1, b) + f(a_2, b)$;
(2) $f(a, b_1 + b_2) = f(a, b_1) + f(a, b_2)$;
(3) $f(ar, b) = f(a, rb)$.

Remark 2.1.1. Pay attention to directions. We can write $f(ar, b)$ but not $f(ra, b)$. Also, it would not make sense to require that $f(ar, b) = rf(a, b)$ because H is not an R-module; it is merely an abelian group.

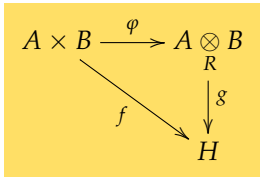
Our goal is to construct an abelian group, called the **tensor product** of A and B over R,

$$A \underset{R}{\otimes} B,$$

which we will also write as $A \otimes_R B$ for typographical reasons, together with an R-biadditive function

$$\varphi\colon A \times B \to A \underset{R}{\otimes} B,$$

such that for any R-biadditive map $f\colon A \times B \to H$, there is a unique group homomorphism $g : A \otimes_R B \to H$, such that the following diagram commutes:[4]



Moreover, if S is another ring and $A \in {}_{\mathbf{S}}\mathbf{Mod_R}$ (see the exercises) then $A \otimes_R B \in_{\mathbf{S}} \mathbf{Mod}$ and we get a functor

$$A \otimes_R (-)\colon {}_{\mathbf{R}}\mathbf{Mod} \to {}_{\mathbf{S}}\mathbf{Mod}.$$

On the other hand, we have the functor

$$\mathrm{Hom}_S(A, (-))\colon {}_{\mathbf{S}}\mathbf{Mod} \to {}_{\mathbf{R}}\mathbf{Mod},$$

and we will prove that

$$(A \otimes_R (-), \mathrm{Hom}_S(A, -))$$

is an adjoint pair. Meaning, for all $B \in {}_{\mathbf{R}}\mathbf{Mod}$, $C \in {}_{\mathbf{S}}\mathbf{Mod}$ we have a natural isomorphism

$$\varphi_{B,C}\colon \mathrm{Hom}_S(A \otimes_R B, C) \cong \mathrm{Hom}_R(B, \mathrm{Hom}_S(A, C)).$$

---

[3] We allow ourselves the more relaxed notation $A \in \mathbf{Mod_R}$ for $A \in ob(\mathbf{Mod_R})$.

[4] We say that $A \otimes_R B$ has a *universal property*. Note that we may define a category whose objects are R-biadditive maps $f\colon A \times B \to H$ and a morphism from $f_1\colon A \times B \to H_1$ to $f_2\colon A \times B \to H_2$ is a homomorphism $g : H_1 \to H_2$ such that $g \circ f_1 = f_2$. In this language, the universal property is asserting that $\varphi\colon A \times B \to A \underset{R}{\otimes} B$ is an *initial object* for this category.

2.2. **Construction of the tensor product.** Let R be a ring, and let $A \in \mathbf{Mod_R}$ and $B \in {}_{\mathbf{R}}\mathbf{Mod}$ be modules. We start by forming the free abelian group G on the symbols $(a,b) \in A \times B$:

$$G = \oplus_{(a,b)\in A\times B}\mathbb{Z}\cdot(a,b).$$

We let $N \subset G$ be the subgroup generated by all the expressions

$$\begin{array}{ll} (a_1+a_2,b)-(a_1,b)-(a_2,b) & a_1,a_2\in A, b\in B \\ (a,b_1+b_2)-(a,b_1)-(a,b_2) & a\in A, b_1,b_2\in B \\ (ar,b)-(a,rb) & a\in A, b\in B, r\in R \end{array}$$

We let

$$A \underset{R}{\otimes} B = G/N.$$

The image of the element (a,b) of G is denoted $a\otimes b$. The map

$$\varphi\colon A\times B \to A\underset{R}{\otimes} B, \qquad (a,b)\mapsto a\otimes b,$$

is R-biadditive by construction. The group $A\otimes_R B$ is called the **tensor product** of A and B over R and its elements are called tensors.

Remark 2.2.1. Note that in G the elements $-(a,b) = -1\cdot(a,b)$ and $(-a,b)$ are not equal. We will see that they are equal in $A\otimes_R B$. For a start, note that for any $b\in B$ we have $(0,b) = (0+0,b)$ and so $(0,b)-(0,b)-(0,b)$ belongs to N. Namely, $(0,b)$ belongs to N, and similarly $(a,0)\in N$, which means that in $A\otimes_R B$

$$0 = 0\otimes b = a\otimes 0, \qquad \forall a\in A, b\in B.$$

As $(0,b) = (a+(-a),b)$ we get that $(0,b)-(a,b)-(-a,b)$ belongs to N which means that in $A\otimes_R B$ we have $0 = (0,b) = (a,b)+(-a,b)$. Namely,

$$(-a)\otimes b = -(a\otimes b), \quad a\otimes(-b) = -(a\otimes b).$$

By similar considerations, for any $n\in\mathbb{Z}$ we have

$$n(a\otimes b) = na\otimes b = a\otimes nb.$$

This implies that we can write the general element of $A\otimes_R B$ as $\sum_i a_i\otimes b_i$. Note though that this presentation is not unique. For example, as we have seen, for any a,b we have $a\otimes 0 = 0\otimes b$. We call an element $a\otimes b$ of $A\otimes_R B$ a **pure tensor**; a general element is a sum of pure tensors but very rarely a pure tensor itself.

Proposition 2.2.2. *$A\otimes_R B$ has the required universal property.*

Proof. Given an R-biadditive into an abelian group

$$f\colon A\times B\to H,$$

define

$$\tilde{g}\colon G\to H, \quad \tilde{g}(\sum_i n_i(a_i,b_i)) = \sum_i n_i f(a_i,b_i),$$

which is a well-defined homomorphism of groups. For every generator z of N we have $\tilde{g}(z) = 0$ and so $N\subseteq \mathrm{Ker}(\tilde{g})$. By the first isomorphism theorem, $\tilde{g}$ induces a well-defined homomorphism of abelian groups

$$g\colon A\otimes_R B\to H, \quad g(\sum_i a_i\otimes b_i) = \sum_i f(a_i,b_i).$$

Note that $g \circ \varphi = f$ by construction, and this property determines g uniquely because g is determined by its values on the pure tensors $a \otimes b$, $a \in A, b \in B$. □

2.3. **Properties of the tensor product.** The following proposition shows functoriality of the tensor product construction. Although a very technical statement, it will turn out to be very important.

Proposition 2.3.1. (Functoriality of tensor products) *Let $f \in \mathrm{Hom}_R(A, A'), g \in \mathrm{Hom}_R(B, B')$. There is a group homomorphism*

$$f \otimes g \colon A \otimes_R B \to A' \otimes_R B',$$

such that

$$(f \otimes g)(a \otimes b) = f(a) \otimes g(b).$$

Proof. The map

$$f \times g \colon A \times B \to A' \otimes_R B', \qquad (f \times g)(a, b) = f(a) \otimes g(b),$$

is an R-biadditive map. There is therefore a group homomorphism, that we call $f \otimes g$, from $A \otimes_R B \to A' \otimes_R B'$ making the following diagram commutative:

$$\begin{array}{ccc} A \times B & \xrightarrow{\varphi} & A \otimes_R B \\ & \searrow^{f \times g} & \downarrow{\scriptstyle f \otimes g} \\ & & A' \otimes_R B' \end{array}$$

By construction, $(f \otimes g)(a \otimes b) = f(a) \otimes g(b)$. □

Corollary 2.3.2. *If $A \cong A', B \cong B'$ then $A \otimes_R B \cong A' \otimes_R B'$.*

Proposition 2.3.3. *We have the following properties of tensor products:*

1. *$A \otimes_R R = A$, $R \otimes_R B \cong B$.*
2. *$(\oplus_{i \in I} A_i) \otimes_R B \cong \oplus_i (A_i \otimes_R B)$, $A \otimes (\oplus_{i \in I} B_i) \cong \oplus_{i \in I} A \otimes_R B_i$.*
3. *If R is commutative (and so we can view any left R-module as a right R-module and vice-versa) then $A \otimes_R B \cong B \otimes_R A$.*

Proof. The proofs are incredibly dull, so we refer to one of the references for the proofs (or, better yet, write your own proofs!). We only prove that $A \otimes_R R = A$ and we do that mainly to make a point. It is tempting to define

$$g \colon A \otimes_R R \to A$$

by $g(\sum_i a_i \otimes r_i) = \sum_i a_i r_i$, but we can't do that yet because, due to the non-uniqueness of expressing tensors as $\sum_i a_i \otimes r_i$, it is not clear that this is well-defined. So, one starts by defining a map $f \colon A \times R \to A$ by $f(a, r) = ar$. This map is clearly a well-defined R-biadditive map and by the property of the tensor product we get a group homomorphism g as we want.

On the other hand, we can define a homomorphism of groups

$$\iota \colon A \to A \otimes_R R, \quad \iota(a) = a \otimes 1.$$

We claim that ι is surjective. Indeed, any element of $A \otimes_R R$ is of the form $\sum_i a_i \otimes r_i = \sum_i a_i r_i \otimes 1 = (\sum_i a_i r_i) \otimes 1$. Since ι is surjective and $g \circ \iota = Id_A$, it follows that both g and ι are isomorphisms. □

Example 2.3.4. Let k be a field and V, W finite dimensional vector spaces over k. Then, *canonically*,

$$\mathrm{Hom}_k(V,W) \cong V^* \otimes_k W.$$

Proof. We remark that both sides are k-vector spaces. Indeed, as k is commutative, V and V^* are k-bimodules and the fact that $V^* \otimes_k W$ is a k-module follows from a more general statement – see Theorem 2.6.1 below. That said, the k-module structure is simple: $\lambda \sum_i \phi_i \otimes w_i = \sum_i \lambda\phi_i \otimes w_i$.

Let $\{v_i\}_{i=1}^d$ be a basis for V and $\{v_i^*\}$ the dual basis of V^*. Then,

$$V^* \otimes_k W = (\oplus_{i=1}^d k \cdot v_i^*) \otimes_k W \cong \oplus_{i=1}^d k \otimes_k W \cong \oplus_{i=1}^d W.$$

Therefore,

$$\dim_k(V^* \otimes_k W) = \dim(V) \cdot \dim(W) = \dim(\mathrm{Hom}_k(V,W)).$$

That proves that $V^* \otimes_k W$ and $\mathrm{Hom}_k(V,W)$ are isomorphic as k-vector spaces, but our goal is to construct a *canonical* isomorphism and it will suffice to construct a canonical surjective k-linear map $V^* \otimes_k W \to \mathrm{Hom}_k(V,W)$. As usual, we start by constructive a bi-additive map. Consider

$$V^* \times W \to \mathrm{Hom}_k(V,W), \qquad (\phi, w) \mapsto \{v \mapsto \phi(v) \cdot w\}.$$

It is easy to verify that $v \mapsto \phi(v) \cdot w$ is an element of $\mathrm{Hom}_k(V,W)$ and that the map we constructed is k-biadditive. Thus, we get an induced map

$$V^* \otimes W \to \mathrm{Hom}_k(V,W), \quad \sum_i \phi_i \otimes w_i \mapsto \{v \mapsto \sum_i \phi_i(v) \cdot w_i\}.$$

This is a surjective map. Indeed, given a linear map $T \in \mathrm{Hom}_k(V,W)$ denote $u_j = T(v_j)$ and consider the element $\sum_i v_i^* \otimes u_i \in V^* \otimes_k W$. The linear map associated to it is precisely T because both maps send v_j to u_j. □

2.4. **Tensor product of algebras.** Let R be a commutative ring. A ring A is called an R-**algebra** if we are given a homomorphism of rings

$$i_A \colon R \to A$$

whose image is contained in the centre $Z(A)$ of A. That is,

$$i_A(r) \cdot a = a \cdot i_A(r), \quad \forall r \in R, a \in A.$$

Note that there is no requirement that i_A is injective.

For example, any ring is a $\mathbb{Z}$-algebra, via

$$i_A(n) = n \cdot 1_A, \quad n \in \mathbb{Z}.$$

(Meaning, if n is positive $n \cdot 1_A = 1_A + \cdots + 1_A$ n-times; if n is negative $n \cdot 1_A$ is defined as $-(-n) \cdot 1_A$ and $0 \cdot 1_A = 0$.)

Any R-algebra is an R-module: $r \cdot a := \iota_A(r)a$; in fact, it is an R-bimodule, where we let $a \cdot r = \iota_A(r)a$; this uses the commutativity of R.

Theorem 2.4.1. *If A, B are R-algebras so is $A \otimes_R B$ and we have*

$$a \otimes b \cdot a' \otimes b' = aa' \otimes bb'.$$

Proof. We will use the functoriality of tensor products (Proposition 2.3.1). For $s \in A$, $t \in B$, the maps

$$[s]\colon A \to A, \quad [s](a) = sa; \qquad [t]\colon B \to B, \quad [t](b) = tb \ \ (sic!),$$

are homomorphisms of R-modules as

$$[s](a \cdot r) = s(a\iota_A(r)) = (sa)\iota_A(r) = ([s](a)) \cdot r, \quad [t](r \cdot b) = t(\iota_B(r)b) = \iota_B(r)[t](b) = r \cdot [t](b).$$

We therefore get a group homomorphism

$$[s] \otimes [t]\colon A \otimes_R B \to A \otimes_R B, \quad ([s] \otimes [t])(\sum_i a_i \otimes b_i) = \sum_i sa_i \otimes tb_i.$$

In fact (!), we get an R-biadditive map

$$A \times B \to \mathrm{End}_{\mathbf{Gps}}(A \otimes_R B), \quad (s,t) \mapsto [s] \otimes [t].$$

Thus, we get a well-defined homomorphism of groups

$$A \otimes_R B \to \mathrm{End}_{\mathbf{Gps}}(A \otimes_R B), \quad \sum_i s_i \otimes t_i \mapsto \sum_i [s_i] \otimes [t_i].$$

Define then,

$$(\sum s_i \otimes t_i)(\sum a_i \otimes b_i) = (\sum_i [s_i] \otimes [t_i])(\sum a_i \otimes b_i) = \sum_{ij} s_i a_j \otimes t_i b_j.$$

This is well-defined and is linear in each of the terms, $\sum s_i \otimes t_i$ and $\sum a_i \otimes b_i$. The ring axioms now become an easy verification, especially as we have a formula for the multiplication. Finally, make $A \otimes_R B$ into an R-algebra by letting

$$\iota(r) = \iota_A(r) \otimes 1 = 1 \otimes \iota_B(r).$$

□

2.5. **Further examples of tensor products.**

(1) Let R be a commutative ring, $I \lhd R$ and M a left R-module then

$$R/I \otimes_R M \cong M/IM.$$

In particular, if $J \lhd R$ then

$$R/I \otimes_R R/J \cong R/(I+J).$$

Proof. We have a map $R/I \times M \to M/IM$ given by $(\bar{r}, m) \mapsto \overline{rm}$, where $\bar{r}$ denotes $r \pmod I$ and $\bar{m}$ denotes $m \pmod{IM}$. It is not hard to check that this map is indeed a well-defined R-biadditive map. It induces a well-defined homomorphism

$$R/I \otimes M \to M/IM, \quad \bar{r} \otimes m \mapsto \overline{rm}.$$

We can also define a map $M \to R/I \otimes_R M$ that takes m to $\bar{1} \otimes m$. This is again a group homomorphism and if $i \in I, m \in M$ then $im \mapsto \bar{1} \otimes im = \bar{i} \otimes m = \bar{0} \otimes m = 0$. Thus, we have an induced group homomorphism $M/IM \to R/I \otimes_R M, \bar{m} \mapsto \bar{1} \otimes m$. On the generators of $R/I \otimes M$, which are elements of the form $\bar{r} \otimes m$, and on elements of M/IM, these maps are inverses of each other. Thus, we get an isomorphism. Finally, this is an isomorphism of R-modules, and not just abelian groups, because both maps are R-module homomorphisms: $r_1 \cdot \bar{r}_2 \otimes m = \overline{r_1 r_2} \otimes m \mapsto \overline{r_1 r_2 m} = r_1 \cdot \overline{r_2 m}$. And for the other map, $rm \mapsto \bar{1} \otimes rm = \bar{r} \otimes m = r \cdot (\bar{1} \otimes m)$.

Applying this to the module R/J we get that $R/I \otimes_R R/J \cong (R/J)/(I \cdot R/J)$. As $I \cdot R/J = (I+J)/J$ we get that $R/I \otimes_R R/J \cong (R/J)/((I+J)/J) \cong R/(I+J)$. □

To illustrate this, take $\mathbb{Z}$ and the ideals $n\mathbb{Z}, m\mathbb{Z}$ to find that

$$\mathbb{Z}/m\mathbb{Z} \otimes \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/\gcd(m,n)\mathbb{Z}.$$

This shows that the tensor product of non-zero modules can be zero (e.g., $m = 2, n = 3$).

(2) Let k be a commutative ring. Then

$$k[x_1,\dots,x_m] \otimes_k k[y_1,\dots,y_n] \cong k[x_1,\dots,x_m,y_1,\dots,y_n].$$

Proof. We will write $f(\underline{x})$ for $f(x_1,\dots,x_m)$ and similarly $f(\underline{y})$. We have a k-biadditive map

$$k[x_1,\dots,x_m] \times_k k[y_1,\dots,y_n] \to k[x_1,\dots,x_m,y_1,\dots,y_n], \quad (f(\underline{x}),g(\underline{y})) \mapsto f(\underline{x})g(\underline{y}),$$

that induces a homomorphism of k-modules

$$k[x_1,\dots,x_m] \otimes_k k[y_1,\dots,y_n] \ \to \ k[x_1,\dots,x_m,y_1,\dots,y_n].$$

In fact, referring to the ring structure on the left hand side, this is a homomorphism of rings. Note that $x_i \otimes 1 \mapsto x_i, 1 \otimes y_j \mapsto y_j$. It thus follows that this is a surjective homomorphism.

We can also define a ring homomorphism,

$$k[x_1,\dots,x_m,y_1,\dots,y_n] \to k[x_1,\dots,x_m] \otimes_k k[y_1,\dots,y_n],$$

as the unique extension to a homomorphism of k-algebras of the function taking x_i (resp. y_j) to $x_i \otimes 1$ (resp. $1 \otimes y_j$). In particular, elements of the form $f(\underline{x}) \otimes g(\underline{y})$ are in the image of this map (this is the image of $f(\underline{x})g(\underline{y})$). Thus, this ring homomorphism is also surjective. If we follow it by the first map, we get the identity. Indeed, it is enough to verify that on generators and we have

$$x_i \mapsto x_i \otimes 1 \mapsto x_i, \quad y_j \mapsto 1 \otimes y_j \mapsto y_j.$$

It follows that both maps are isomorphisms of k-algebras. □

In terms of algebraic geometry, a subject we will touch upon later, this means that if we let $\mathbb{A}^m_k$ be the affine space of dimension m over k (it is associated to the k-algebra $k[x_1,\dots,x_m]$) then

$$\mathbb{A}^m_k \times \mathbb{A}^n_k \cong \mathbb{A}^{m+n}_k.$$

Similarly, the following example, left as an exercise, has an interpretation in terms of subvarieties of affine spaces.

(3) Let k be a commutative ring and $I \lhd k[\underline{x}], J \lhd k[\underline{y}]$. Then

$$k[\underline{x}]/I \underset{k}{\otimes} k[\underline{y}]/J \cong k[\underline{x},\underline{y}]/(\langle I\rangle + \langle J\rangle), \tag{2}$$

where $\langle I\rangle$ is the ideal generated by I in $k[\underline{x},\underline{y}]$. That is, the ideal $Ik[\underline{x},\underline{y}]$.

To illustrate this, we show that we have an isomorphism of rings

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \times \mathbb{C}.$$

The left hand side may be viewed as

$$\mathbb{R}[x]/(x^2+1) \otimes \mathbb{R}[y]/(y^2+1) \cong \mathbb{R}[x,y]/(x^2+1,y^2+1) \cong (\mathbb{R}[x]/(x^2+1))[y]/(y^2+1),$$

which is isomorphic to $\mathbb{C}[y]/(y^2+1)$. Now, applying the Chinese remainder theorem to the ideal $(y^2+1) = (y-i)(y+i)$, we get that $\mathbb{C}[y]/(y^2+1) \cong \mathbb{C} \times \mathbb{C}$, by the map taking y to $(i,-i)$. If one traces through the definitions, one find that the isomorphism is given by

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \times \mathbb{C}, \quad z_1 \otimes z_2 \mapsto (z_1z_2, z_1\bar{z}_2).$$

This can be generalized considerably. Let L/K be a finite Galois extension of fields and $G = \mathrm{Gal}(L/K)$. Then

$$L \otimes_K L \cong \prod_{\sigma \in G} L, \quad \ell \otimes \lambda \mapsto (\ell \cdot \sigma(\lambda))_{\sigma \in G}.$$

Finally, we also remark that the isomorphism (2) is true also when the number of variables is zero and thus the statement here includes the statement $R/I \otimes_R R/J \cong R/(I+J)$ that we showed previouly.

2.6. **The adjoint property for $\otimes$ and Hom.** Let R, S be rings.

Theorem 2.6.1. *Let $A \in {}_{\mathbf{S}}\mathbf{Mod}_{\mathbf{R}}$. The functors*

$$A \underset{R}{\otimes} (-) \colon {}_{\mathbf{R}}\mathbf{Mod} \to {}_{\mathbf{S}}\mathbf{Mod}, \qquad \mathrm{Hom}_S(A,-) \colon {}_{\mathbf{S}}\mathbf{Mod} \to {}_{\mathbf{R}}\mathbf{Mod}$$

are an adjoint pair of functors

$$\left(A \underset{R}{\otimes} (-),\ \mathrm{Hom}_S(A,-)\right)$$

Namely, there are natural isomorphisms $\varphi_{B,C}$ for $B \in {}_{\mathbf{R}}\mathbf{Mod}, C \in {}_{\mathbf{S}}\mathbf{Mod}$

$$\mathrm{Hom}_S(A \otimes_R B, C) \xrightarrow[\cong]{\varphi_{B,C}} \mathrm{Hom}_R(B, \mathrm{Hom}_S(A,C)).$$

Similarly, if $B \in {}_{\mathbf{R}}\mathbf{Mod}_{\mathbf{S}}$, $((-) \otimes_R B, \mathrm{Hom}_S(B,-))$ is an adjoint pair.

In the following proof we leave many routine verifications to the reader. This is an exercise one should do.

Proof. We know that tensor and Hom are functors into abelian groups. We first use that A is a bi-module to show that the tensor and Hom are actually modules.

- $A \otimes_R B$ is an S-module. To see that, let $s \in S$ and consider the map $[s] \colon A \to A, [s](a) = sa$. This is a homomorphism of R-modules and so we get a well defined homomorphism $[s] \otimes Id \colon A \otimes_R B \to A \otimes_R B$. We define then the S-module structure by

  $$s \cdot a \otimes b = ([s] \otimes Id)(a \otimes b) = sa \otimes b.$$

  It is well defined and $s \cdot \sum_i a_i \otimes b_i = \sum sa_i \otimes b_i$, which makes the verification of the module axioms easy.
- $\mathrm{Hom}_S(A,C))$ is an R-module. We define the module structure by

  $$(r \cdot f)(a) := f(ar), \qquad f \in \mathrm{Hom}_S(A,C), a \in A, r \in R.$$

  Once more the verification of the module axioms is straightforward.

We now construction a map between the modules. For $f \in \mathrm{Hom}_S(A \otimes_R B, C)$ define a function $\phi \in \mathrm{Hom}_R(B, \mathrm{Hom}_S(A,C))$ by

$$b \mapsto \phi_b \in \mathrm{Hom}_S(A,C), \qquad \phi_b(a) = f(a \otimes b).$$

There are few points to verify:

- $\phi_b \in \mathrm{Hom}_S(A,C)$. Namely, that $\phi_b(a)$ is linear in a and that $\phi_b(sa) = s\phi_b(a)$. Let's verify the later: $\phi_b(sa) = f(sa \otimes b) = sf(a \otimes b) = s\phi_b(a)$.

- That $b \mapsto \phi_b$ is a homomorphism of R-modules. That amounts to $\phi_{b_1+b_2} = \phi_{b_1} + \phi_{b_2}$ as maps, which one verifies by evaluating both sides at $a \in A$. And, $\phi_{rb} = r\phi_b$, which again one verifies by evaluation at $a \in A$.
- Finally, one needs to check that $f \mapsto \phi$ is a homomorphism of groups. Again, a straightforward verification.

We now construct a map in the opposite direction. Let $\phi \in \mathrm{Hom}_R(B, \mathrm{Hom}_S(A,C))$, say $b \mapsto \phi_b \in \mathrm{Hom}_S(A,C)$. We define $f \in \mathrm{Hom}_S(A \otimes_R B, C)$ as follows. We begin with the function

$$A \times B \to C, \quad (a,b) \mapsto \phi_b(a).$$

- This is an R-biadditive map as, for example, $(ar,b) \mapsto \phi_b(ar) \stackrel{!}{=} (r\phi_b)(a) \stackrel{!}{=} \phi_{rb}(a)$ which is the image of (a,rb). (Make sure you understand why each of the equalities holds true!)
- We thus get a homomorphism of groups

$$f\colon A \otimes_R B \to C, \qquad f(a \otimes b) = \phi_b(a).$$

  In fact, $f \in \mathrm{Hom}_S(A \otimes_R B, C)$ – it respects the S-module structure.
- This construction is inverse to the previous. Once verified, it follows that the construction $\phi \mapsto f$ is also a homomorphism of groups.

Thus, at this point, we have defined an isomorphism of groups,

$$\mathrm{Hom}_S(A \otimes_R B, C) \xrightarrow{\varphi_{B,C}} \mathrm{Hom}_R(B, \mathrm{Hom}_S(A,C)).$$

Remark 2.6.2. Before continuing with the proof, we note that what we have achieved so far, combined with Example 2.3.4, implies that for finite dimensional k-vectors spaces (take $R = C = k$)

$$(V \otimes_k W)^* \cong \mathrm{Hom}_k(W, V^*) \cong W^* \otimes_k V^* \cong V^* \otimes_k W^*.$$

Note also that by the tensor product property

$$(V \otimes_k W)^* = \text{bilinear forms } V \times W \to k.$$

Back to the proof. One needs now to verify that $\varphi_{B,C}$ is natural in B and C. So, for example, we need to verify that for $\gamma \in \mathrm{Hom}_S(C, C')$ the following diagram commutes:

- The following diagram commutes

$$\begin{array}{ccc} \mathrm{Hom}_S(A \otimes_R B, C) & \xrightarrow{\varphi_{B,C}} & \mathrm{Hom}_R(B, \mathrm{Hom}_S(A,C)) \\ \downarrow{\scriptstyle \gamma\circ(-)} & & \downarrow{\scriptstyle \gamma\circ(-)} \\ \mathrm{Hom}_S(A \otimes_R B, C') & \xrightarrow{\varphi_{B,C'}} & \mathrm{Hom}_R(B, \mathrm{Hom}_S(A,C)) \end{array}$$

  Indeed, $f \in \mathrm{Hom}_S(A \otimes_R B, C)$ goes to $\varphi_{B,C}(f) = \{b \mapsto \phi_b, \phi_b(a) = f(a \otimes b)\}$. Under the vertical right arrow, this goes to $b \mapsto \gamma \circ \phi_b, (\gamma \circ \phi_b)(a) = \gamma(f(a \otimes b))$. From the other side, $f \mapsto \gamma \circ f$ that then maps to the function that takes b to the homomorphism $A \to C'$ taking a to $(\gamma \circ f)(a \otimes b)$. So, we are good.
- One verifies naturality relative to $B \to B'$ in a similar way.

□

2.6.1. *Application: Frobenius reciprocity.* Let $H < G$ be finite groups and let k be a field. Any k-representation (ρ, W) of G,

$$W = k - \text{vector space}, \quad \rho \colon G \to \mathrm{GL}_k(W),$$

induces a k-representation of H, $(\rho|_H, W)$, where

$$\rho|_H \colon H \to \mathrm{GL}_k(W), \quad \rho|_H(h) = \rho(h).$$

Very simple! Another way to view the situation is to think of W as a $k[G]$-module and note that $k[H] \subset k[G]$ and thus W becomes a $k[H]$-module. This provides us with a **restriction functor**

$$\mathrm{Res}^G_H \colon {}_{\mathbf{k[G]}}\mathbf{Mod} \to {}_{\mathbf{k[H]}}\mathbf{Mod}.$$

On the other hand $k[G]$ is a bimodule in ${}_{\mathbf{k[G]}}\mathbf{Mod}_{\mathbf{k[H]}}$ and so we get the **induction functor**

$$\mathrm{Ind}^G_H \colon {}_{\mathbf{k[H]}}\mathbf{Mod} \to {}_{\mathbf{k[G]}}\mathbf{Mod}, \qquad \mathrm{Ind}^G_H(V) = k[G] \underset{k[H]}{\otimes} V.$$

It should be remarked that this is a rather remarkable construction: a representation of a subgroup induces a representation of the group G and in a rather non-trivial fashion. Even when $H = \{1\}$ and V is the trivial one-dimensional representation of H, the induced representation is the regular representation $k[G]$, which is a very rich representation of G (it contains any irreducible representation of G, for example). Note also that

$$\dim_k(\mathrm{Res}^G_H(W)) = \dim_k(W), \quad \dim_k(\mathrm{Ind}^G_H(V)) = [G : H] \dim_k(V).$$

Now, using the tensor-Hom adjoint property, we find that

$$\mathrm{Hom}_{k[G]}(\mathrm{Ind}^G_H(V), W) = \mathrm{Hom}_{k[H]}(V, \mathrm{Hom}_{k[G]}(k[G], W)) = \mathrm{Hom}_{k[H]}(V, W) = \mathrm{Hom}_{k[H]}(V, \mathrm{Res}^G_H W).$$

Theorem 2.6.3. *Let k be a field and $H < G$ finite groups.*

Frobenius Reciprocity: $\mathrm{Hom}_{k[G]}(\mathrm{Ind}^G_H(V), W) = \mathrm{Hom}_{k[H]}(V, \mathrm{Res}^G_H W)$

We will later see a more precise and easily applicable version of Frobenius reciprocity using characters. But it's worth mentioning that this version works over any field, including when $\mathrm{char}(k) | \sharp G$ where the group ring is not semi-simple (and thus the theory of characters, as we are going to develop it, does not apply).

We illustrate the power of this theorem. Suppose that V is an irreducible representation. Namely, suppose that it is a simple $k[H]$-module, or, equivalently, that any invariant subspace of V is either $\{0\}$ or V. When is $\mathrm{Ind}^G_H V$ irreducible? Suppose that $k = \mathbb{C}$. We shall see that a representation is irreducible if and only if it has very few endomorphisms. In particular, $\mathrm{Ind}^G_H V$ is irreducible if and only if

$$1 = \dim_{\mathbb{C}} \mathrm{Hom}_{k[G]}(\mathrm{Ind}^G_H V, \mathrm{Ind}^G_H V) = \dim_{\mathbb{C}} \mathrm{Hom}_{k[H]}(V, \mathrm{Res}^G_H \mathrm{Ind}^G_H V).$$

This happens if and only if V appears with multiplicity 1 in $\mathrm{Res}^G_H \mathrm{Ind}^G_H V$ and that is something that is often easy to check in practice.

2.7. **Tensor products over a commutative ring.** Let R be a commutative ring and let A, B be R-modules. We can then think of A, B as R-bimodules, where $ra = ar, rb = br, \forall a \in A, b \in B, r \in R$. Then $A \otimes_R B$ is an R-module as well, where $r \cdot a \otimes b = ra \otimes b = a \otimes rb$. In particular, if C is another R module, we can make more complicated constructions and one can prove that

$$(A \otimes_R B) \otimes_R C \cong A \otimes_R (B \otimes_R C).$$

Let H be an R-module and let $f\colon A\times B\to H$ be an R-bilinear map of R-modules from $A\times B$ to H. Meaning, besides being R-biadditive it also satisfies $f(ra,b)=rf(a,b)$ (and hence also $f(a,rb)=rf(a,b)$). Thinking of f just as R-biadditive, we get a group homomorphism

$$g\colon A\otimes_R B\to H,\quad g(\sum a_i\otimes b_i)=\sum f(a_i,b_i).$$

This homomorphism is automatically an R-module homomorphism as

$$g(r(\sum a_i\otimes b_i))=g(\sum ra_i\otimes b_i)=\sum f(ra_i,b_i)=\sum rf(a_i,b_i)=r\sum f(a_i,b_i)=rg(\sum a_i\otimes b_i).$$

It follows that $A\otimes_R B$ has a universal property in the category of R-modules, namely relative to R-bilinear maps $A\times B\to H$ into R-modules H.

Finally, we remark that when R is commutative one can construct $A\otimes_R B$ a different way; it has the advantage that from the very start it is clear that $A\otimes_R B$ is an R-module. We can start with the free R-module $G'=\oplus_{(a,b)\in A\times B}R\cdot(a,b)$ and take the quotient by the R-submodule N' generated by all expressions of the form $(a_1+a_2,b)-(a_1,b)-(a_2,b),(a,b_1+b_2)-(a,b_1)-(a,b_2),(ra,b)-(a,rb)$, and $r(a,b)-(ra,b)$. It easy to modify the proof we gave for the tensor product G/N to show that G'/N' has the universal property for R-bilinear maps into R-modules, which gives a unique isomorphism $G/N\cong G'/N'$ (initially of abelian groups, but then one checks that also as R-modules).

3. Localization

Throughout our discussion of localization R is a commutative ring. A set $S\subseteq R$ is called a **multiplicative set** if

$$1\in S,\quad x,y\in S\Longrightarrow xy\in S.$$

Example 3.0.1. (1) $S=\{1,f,f^2,\cdots\}$ for some $f\in R$.
(2) $S=R\setminus\mathfrak{p}$, where $\mathfrak{p}$ is a prime ideal of R.
(3) If R is an integral domain, (0) is a prime ideal and so we may take $S=R-\{0\}$.

Our goals are the following.

(1) Construct a ring $R[S^{-1}]$ (R localized at S) with a ring homomorphism
$$R\to R[S^{-1}],$$
having a certain universal property.
(2) Construct a localization functor
$$_{\mathbf{R}}\mathbf{Mod}\to{}_{\mathbf{R}[\mathbf{S}^{-1}]}\mathbf{Mod},\quad M\mapsto M[S^{-1}].$$
(3) Compare properties of R and M with properties of their localizations.

3.1. **Construction of the localization.** Let $M\in_{\mathbf{R}}\mathbf{Mod}$ and $S\subset R$ a multiplicative set. The case $M=R$ is allowed and is particularly important. We consider formal fractions

$$\frac{m}{s},\quad m\in M,s\in S.$$

We define a relation on fractions by

$$\frac{m_1}{s_1}\sim\frac{m_2}{s_2}\text{ if }\exists s\in S\text{ such that }s(s_2m_1-s_1m_2)=0.$$

We leave it as an exercise to check that this is an equivalence relation. We denote $M[S^{-1}]$ the set of equivalence classes of formal fractions m/s. We will usually be cavalier in our notation and not distinguish notationally between a fraction m/s and its equivalence class. We define an operation on $M[S^{-1}]$:

$$\frac{m_1}{s_1} + \frac{m_2}{s_2} = \frac{s_2 m_1 + s_1 m_2}{s_1 s_2}.$$

We check that this is well-defined. If $m_i/s_i \sim n_i/t_i$, namely, $\exists \alpha_i \in S, \alpha_i(t_i m_i - s_i n_i) = 0$, then

$$\frac{s_2 m_1 + s_1 m_2}{s_1 s_2} \sim \frac{t_2 n_1 + t_1 n_2}{t_1 t_2}.$$

Indeed, $\alpha_1\alpha_2 \in S$ and $\alpha_1\alpha_2(t_1t_2(s_2m_1 + s_1m_2) - s_1s_2(t_2n_1 + t_1n_2)) = \alpha_1\alpha_2(t_2s_2(t_1m_1 - s_1n_1) + t_1s_1(t_2m_2 - s_2n_2)) = 0$. It is now straightforward to verify that this operation makes $M[S^{-1}]$ into an abelian group, called the **localization of** M **at** S, whose 0 element is the class of $\frac{0}{1}$.

It is useful to note that

$$0 = \frac{0}{s}, \quad \forall s \in S; \qquad \frac{m}{s} = \frac{s_1 m}{s_1 s}, \quad \forall s, s_1 \in S, m \in M.$$

In the case $M = R$ we can make $R[S^{-1}]$ into a ring by defining

$$\frac{r_1}{s_1} \cdot \frac{r_2}{s_2} = \frac{r_1 r_2}{s_1 s_2}.$$

We leave it as an exercise to check that $R[S^{-1}]$ is a ring and $M[S^{-1}]$ is an $R[S^{-1}]$-module where the module structure is given by

$$\frac{r}{s} \cdot \frac{m}{s_1} = \frac{rm}{ss_1}, \quad r \in R, m \in M,\ s, s_1 \in S.$$

Note also that the map

$$R \to R[S^{-1}], \qquad r \mapsto \frac{r}{1}$$

is a ring homomorphism with kernel

$$\{r \in R : \exists s \in S, sr = 0\}.$$

Thus, if R is an integral domain and $0 \notin S$, the map $R \to R[S^{-1}]$ is injective, but in general it need not be. In addition, the homomorphism $f\colon R \to R[S^{-1}]$ has the following universal property. Given any ring T and a ring homomorphism $g\colon R \to T$, such that every element $g(s), s \in S$, is invertible in T, there is a unique homomorphism $h\colon R[S^{-1}] \to T$ such that $h \circ f = g$.

Example 3.1.1. If we take $R = \mathbb{Z}$ and $S = \{1, 2, 2^2, \dots\}$, we find the ring

$$\mathbb{Z}[\tfrac{1}{2}] = \mathbb{Z}[S^{-1}] = \{\tfrac{n}{2^a} : n \in \mathbb{Z}, a \geq 0\}.$$

If we localize the $\mathbb{Z}$-module $\mathbb{Z}/n\mathbb{Z}$ by S we get the zero module if $n = 2^b$, and a module isomorphic to $\mathbb{Z}/n\mathbb{Z}$ if n is odd. Indeed, if $n = 2^b$, for every $m \in \mathbb{Z}/n\mathbb{Z}$ we have $\frac{m}{1} = \frac{2^b m}{2^b} = \frac{0}{2^b} = 0$. If n is odd, the map

$$\mathbb{Z}/n\mathbb{Z} \mapsto (\mathbb{Z}/n\mathbb{Z})[\tfrac{1}{2}], \quad m \mapsto \frac{m}{1},$$

is an isomorphism of abelian groups: first, it is surjective because given $\frac{m}{2^a}$ there is a 2^b such that $2^b 2^a \equiv 1 \pmod{n}$ and we find that $\frac{m}{2^a} = \frac{2^b m}{2^{a+b}} = \frac{2^b m}{1}$. It is also injective: If $\frac{m}{1} = 0$ then for some $2^a \in S$, $2^a m \equiv 0 \pmod{n}$, but, as 2^a is a unit modulo n, it follows that $m \equiv 0 \pmod{n}$.

More generally, using the Chinese Remainder Theorem, we see that if p is a prime then

$$\mathbb{Z}[\tfrac{1}{p}] = \{\tfrac{n}{p^a} : n \in \mathbb{Z}, a \geq 0\},$$

and for a positive integer $n = p^a m$, where $(p, m) = 1$, we have

$$(\mathbb{Z}/n\mathbb{Z})[\tfrac{1}{p}] \cong \mathbb{Z}/m\mathbb{Z}.$$

If p is a prime there is another localization associated to it, which we denote $\mathbb{Z}_{(p)}$, following our notation $R_{\mathfrak{p}}$ for a prime ideal $\mathfrak{p}$ of a ring R. This is the localization at the elements *not* in the prime ideal (p). Namely,

$$\mathbb{Z}_{(p)} = \{\tfrac{m}{n} : m, n \in \mathbb{Z}, p \nmid n\}.$$

One can check that if n is a positive integer, $n = p^a m, (m, p)$, then

$$(\mathbb{Z}/n\mathbb{Z})_{(p)} \cong \mathbb{Z}/p^a\mathbb{Z}.$$

3.2. **Localization is an exact functor.** We show that localization provides a functor

$$_{\mathbf{R}}\mathbf{Mod} \to {}_{\mathbf{R}[\mathbf{S^{-1}}]}\mathbf{Mod}, \quad M \mapsto M[S^{-1}].$$

We have defined what the functor does to objects. Now given a homomorphism $f\colon M \to N$ of R modules we define

$$f[S^{-1}]\colon M[S^{-1}] \to N[S^{-1}], \qquad f[S^{-1}]\left(\frac{m}{s}\right) = \frac{f(m)}{s}.$$

We will usually write f for $f[S^{-1}]$.

Lemma 3.2.1. *$f[S^{-1}]$ is a well-defined homomorphism of $R[S^{-1}]$-modules.*

Proof. We verify the axioms.

- Well-defined. If $m_1/s_1 \sim m_2/s_2$ there is an $s \in S$ such that $s(s_2m_1 - s_1m_2) = 0$. Apply f to this equality to get $s(s_2f(m_1) - s_1f(m_2)) = 0$, which means that $f(m_1)/s_1 \sim f(m_2)/s_2$, as required.
- Additive. $\frac{m_1}{s_1} + \frac{m_2}{s_2} = \frac{s_2m_1+s_1m_2}{s_1s_2} \overset{f}{\mapsto} \frac{s_2f(m_2)+s_1f(m_2)}{s_1s_2} = \frac{f(m_1)}{s_1} + \frac{f(m_2)}{s_2}$.
- Scalars. $f\left(\frac{r}{s} \cdot \frac{m}{s_1}\right) = f\left(\frac{rm}{ss_1}\right) = \frac{rf(m)}{ss_1} = \frac{r}{s} \cdot \frac{f(m)}{s_1}$.

□

It is clear from the definition that

$$Id_M[S^{-1}] = Id_{M[S^{-1}]}, \quad (g \circ f)[S^{-1}] = g[S^{-1}] \circ f[S^{-1}],$$

and thus localization is a functor.

There is another way to conclude that localization is a functor; we can relate the notion of localizing a module and localizing a ring and deduce functoriality of localization of modules from functoriality of the tensor product.

Lemma 3.2.2. *There is a canonical isomorphism of $R[S^{-1}]$-modules,*

$$R[S^{-1}] \otimes_R M \cong M[S^{-1}].$$

Proof. Once the map and its inverse are defined it is easy to verify that these are homomorphism of $R[S^{-1}]$-modules. The map

$$R[S^{-1}] \otimes_R M \cong M[S^{-1}], \quad (\frac{r}{s}, m) \mapsto \frac{rm}{s}$$

is well defined. If $\frac{r}{s} = \frac{r'}{s'}$ then for some $t \in S$ we have $ts'r = tsr'$ and we compute that $\frac{rm}{s} = \frac{ts'rm}{ts's} = \frac{tsr'm}{ts's} = \frac{r'm}{s'}$. The R-biadditivity is easy to check and we get a well-defined homomorphism of groups

$$R[S^{-1}] \otimes_R M \to M[S^{-1}], \quad \sum \frac{r_i}{s_i} \otimes m_i \mapsto \sum \frac{r_i m_i}{s_i},$$

which is easily checked to be a homomorphism of $R[s^{-1}]$-modules.

To define the inverse map we define $\frac{m}{s} \mapsto \frac{1}{s} \otimes m$ and we need to check this is well defined. If $\frac{m}{s} = \frac{m'}{s'}$ then for some $t \in S$ we have $ts'm = tsm'$. Then $\frac{1}{s} \otimes m = \frac{1}{ss't} \otimes s'tm = \frac{1}{ss't} \otimes stm' = \frac{1}{s'} \otimes m'$. It is easily verified that this map is an inverse to the previous map. □

We next prove that localization is an exact functor. To begin with, a **short exact sequence** (SES) of R-modules is a diagram of R-modules and R-modules homomorphisms

$$0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$$

in which f is injective, g is surjective and $\mathrm{Ker}(g) = \mathrm{Im}(f)$. Equivalently, the kernel of any homomorphism is the image of the previous homomorphism. Here are some simple examples:

- $0 \longrightarrow \mathbb{Z} \xrightarrow{\times n} \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z} \longrightarrow 0$, for any $n > 0$.
- $0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Q} \longrightarrow \mathbb{Q}/\mathbb{Z} \longrightarrow 0$.
- $0 \longrightarrow I \longrightarrow R \longrightarrow R/I \longrightarrow 0$, for any ideal $I \lhd R$.

Proposition 3.2.3. *Localization is an* **exact** *functor: it takes a SES in* $_{\mathbf{R}}\mathbf{Mod}$ *to a SES in* $_{\mathbf{R}[\mathbf{S}^{-1}]}\mathbf{Mod}$.

Proof. Let

$$0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$$

be a SES of R-modules. We need to show that

$$0 \longrightarrow M_1[S^{-1}] \xrightarrow{f[S^{-1}]} M_2[S^{-1}] \xrightarrow{g[S^{-1}]} M_3[S^{-1}] \longrightarrow 0$$

is a SES of $R[S^{-1}]$-modules.

- $f[S^{-1}]$ is injective. Suppose $f(m_1/s_1) = f(m_1)/s_1 = 0$. Then $\exists t \in S$ such that $tf(m_1) = 0$. So $f(tm_1) = 0$ and, as f is injective, $tm_1 = 0$. Therefore, $m_1/s_1 = tm_1/ts_1 = 0/ts_1 = 0$.
- $g[S^{-1}]$ is surjective. Given $m_3/s \in M_3[S^{-1}]$, there is some $m_2 \in M_2$ with $g(m_2) = m_3$ and so $g(m_2/s) = m_3/s$.
- $\mathrm{Im}(f[S^{-1}]) = \mathrm{Ker}(g[S^{-1}])$. Since $g \circ f = 0$ also $0 = (g \circ f)[S^{-1}] = g[S^{-1}] \circ f[S^{-1}]$ and so $\mathrm{Im}(f[S^{-1}]) \subseteq \mathrm{Ker}(g[S^{-1}])$. Let $m_2/s \in \mathrm{Ker}(g[S^{-1}])$, meaning for some $t \in S = tg(m_2) = g(tm_2) = 0$. There is thus $m_1 \in M_1$ such that $f(m_1) = tm_2$ and therefore $f(m_1/ts) = tm_2/ts = m_2/s$, showing that $\mathrm{Im}(f[S^{-1}]) \supseteq \mathrm{Ker}(g[S^{-1}])$.

□

3.3. **Behaviour of ideals under localization.** We will see later, in §5, that to a commutative ring R we can associate a topological space $\mathrm{Spec}(R)$, endowed with a sheaf of functions. The points of this space are the set

$$\{\mathfrak{p} : \mathfrak{p} \text{ is a prime ideal of } R\}.$$

To any multiplicative set S in R we can associate a subset of $\mathrm{Spec}(R)$ consisting of primes that "survive" in the localization $R[S^{-1}]$. That is, it is an open set whose points (certain primes ideals of R) are in natural bijection with primes of $R[S^{-1}]$. This is the the geometric reason for understanding how ideals behave under localization, but localization serves as a useful technique in many algebraic arguments, even without recourse to geometry.

3.3.1. *Extended and contracted ideals.* Let $S \subset R$ be a multiplicative set and let

$$f\colon R \to R[S^{-1}]$$

be the canonical ring homomorphism, $f(r) = r/1$. For $J \lhd R[S^{-1}]$ define the **contracted ideal** J^c,

$$J^c = f^{-1}(J) = \{r \in R : \frac{r}{1} \in J\}.$$

This is an ideal of R by general properties of ring homomorphisms.

In the other direction, for $I \lhd R$ define the **extended ideal** I^e,

$$I^e = I[S^{-1}] \subset R[S^{-1}].$$

As $I \hookrightarrow R$ as R-modules, $I[S^{-1}] \hookrightarrow R[S^{-1}]$ as $R[S^{-1}]$-modules. Namely, I^e is an ideal of $R[S^{-1}]$. Note also that

$$I^e := \{\frac{r}{s} : r \in I, s \in S\},$$

which is precisely the ideal of $R[S^{-1}]$ generated by $f(I)$. Thus, we will also write

$$I^e = \langle f(I) \rangle = IR[S^{-1}].$$

3.3.2. *Prime ideals under localization.*

Theorem 3.3.1. *There is a bijection,*

$$\left\{ \begin{array}{c} \textit{prime ideals of } R \\ \textit{disjoint from } S \end{array} \right\} \quad \longleftrightarrow \quad \left\{\textit{prime ideals of } R[S^{-1}]\right\},$$

under which

$$I \longmapsto I^e,$$

$$J^c \longleftarrow\!\shortmid J.$$

Proof. We already know that J^c is an ideal of R. As $R/J^c \hookrightarrow R[S^{-1}]/J$, R/J^c is an integral domain and so J^c is a prime ideal. It is disjoint from S because if $s \in J^c \cap S$ then $f(s) = \frac{s}{1} \in J$ and so J contains a unit, implying $J = R[S^{-1}]$ and that is a contradiction.

Suppose now that I is a prime ideal of R, disjoint from S. We show that I^e is a prime ideal of $R[S^{-1}]$. Suppose that $\frac{m_1}{s_1}\frac{m_2}{s_2} = \frac{m_1 m_1}{s_1 s_2} \in I^e$. Then, for some $i \in I, s_3 \in S$ we have $\frac{m_1 m_1}{s_1 s_2} = \frac{i}{s_3}$ and that implies that for some $s \in S$

$$s s_3 m_1 m_2 = s s_1 s_2 i \in I.$$

Since I is prime, one of s, s_3, m_1, m_2 belongs to I. As I is disjoint from S, one of m_1, m_2 belongs to I; say m_1. Then $\frac{m_1}{s_1} \in I^e$. To show I^e is prime it only remains to check that $I^e \neq R[S^{-1}]$ and that follows from the property $I^{ec} = I$ that we prove next.

It is clear from the construction that $I^{ec} \subset I$. Let $m \in I^{ec}$ then $\frac{m}{1} \in I^e$ and thus for some $i \in I$, $s \in S$ we have $\frac{m}{1} = \frac{i}{s}$. Thus, for some $s_1 \in S$ we have $s_1 sm = s_1 i \in I$. Using, as above, that I is prime and disjoint from S, we get that $m \in I$.

Finally, we show $J^{ce} = J$. The inclusion $J^{ce} \subset J$ is easy: $f(J^c) \subset J$ by definition and since J is an ideal $\langle f(J^c) \rangle = J^{ce} \subset J$. Let $\frac{m}{s} \in J$ then $\frac{m}{1} \in J$ and so $m \in J^c$. But then $\frac{m}{s} = \frac{1}{s} f(m) \in J^{ce}$ and we get the other inclusion.[5] □

Example 3.3.2. If $\mathfrak{p}$ is a prime ideal of R and $S = R \setminus \mathfrak{p}$, it is customary to denote $R[S^{-1}]$ by $R_{\mathfrak{p}}$ and call it the localization of R at $\mathfrak{p}$ (although we are actually localizing at the complement of $\mathfrak{p}$). This ring is

$$R_{\mathfrak{p}} := \left\{ \frac{r}{s} : r \in R, s \notin \mathfrak{p} \right\}.$$

It follows from the theorem that the prime ideals of $R_{\mathfrak{p}}$ are precisely the localization at S of the prime ideals $\mathfrak{q}$ of R such that $\mathfrak{q} \subset \mathfrak{p}$. Among those ideals there is a unique maximal one, namely $\mathfrak{p}$. Thus, $R_{\mathfrak{p}}$ has a unique maximal ideal (a commutative ring with a unique maximal ideal is called a **local ring**).

Thus, we see that to study prime ideals contain in $\mathfrak{p}$ we may pass to the local ring $R_{\mathfrak{p}}$, while to study prime ideals containing $\mathfrak{p}$ we may pass to the integral domain $R/\mathfrak{p}$.

Example 3.3.3. Let R be an integral domain and $S = R - \{0\}$. In this case we denote

$$\mathrm{Frac}(R) = R[S^{-1}] = R_{(0)} = \left\{ \frac{r}{s} : r, s \in R, s \neq 0 \right\},$$

and call it the **fraction field** of R. For example, $\mathbb{Q}$ is obtained this way from $\mathbb{Z}$. It follows from the universal property of localizations that this field embeds into any field in which R embeds. In this sense it is the smallest field containing R.

3.4. **Local properties.** Let R be a commutative ring and $\mathfrak{p}$ a prime ideal of R. Let $S = R - \mathfrak{p}$. We introduced above the notation $R_{\mathfrak{p}}$ for $R[S^{-1}]$ and we now extend it to modules and denote

$$M_{\mathfrak{p}} = M[S^{-1}].$$

A property $\mathcal{P}$ of modules is called a **local property** if

$$M \in {}_{\mathbf{R}}\mathbf{Mod} \text{ has } \mathcal{P} \iff M_{\mathfrak{p}} \in {}_{\mathbf{R}_{\mathfrak{p}}}\mathbf{Mod} \text{ has } \mathcal{P}, \ \forall \mathfrak{p} \lhd R \text{ prime}.$$

Proposition 3.4.1. (Being zero is a local property) *Let M be an R-module. Then $M = \{0\}$ if and only if $M_{\mathfrak{p}} = \{0\}$ for all prime ideals $\mathfrak{p}$ of R.*

Proof. If $M = 0$ then $M[S^{-1}] = \{0\}$ for any multiplicative set $S \subset R$ because all the elements of $M[S^{-1}]$ are of the form $\frac{0}{s}$ and they are all the same zero element of $M[S^{-1}]$.

For the converse, suppose that $M \neq 0$ and take a non-zero element $m \in M$. Then

$$1 \notin \mathrm{Ann(m)} := \{r \in R : rm = 0\},$$

[5] Note that the argument for $I^{ec} = I$ used the primality of I, and indeed it could fail for non-prime ideals, but the argument for $J^{ce} = J$ did not and so the statement is true for arbitrary ideals of $R[S^{-1}]$.

the annihilator ideal of m. Thus, there is some maximal ideal $\mathfrak{p} \supseteq \mathrm{Ann}(m)$. If $\frac{m}{1} = 0$ in $M_\mathfrak{p}$ then for some $s \notin \mathfrak{p}$ we have $sm = 0$. Then, on the one hand $s \in \mathrm{Ann(m)}$ and on the other hand $s \notin \mathfrak{p}$, and so $s \notin \mathrm{Ann(m)}$. Contradiction. It follows that for some maximal ideal $\mathfrak{p}$, $M_\mathfrak{p} \neq 0$. □

Remark 3.4.2. The proof shows that $M = \{0\}$ if and only if $M_\mathfrak{p} = \{0\}$ for all maximal ideals $\mathfrak{p}$ of R.

Proposition 3.4.3. (A complex being exact is a local property) *Let*

$$0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0 \tag{3}$$

be a complex of R modules (meaning, the image of any homomorphism is contained in the kernel of the next, which amounts to $\mathrm{Im}(f) \subset \ker(g)$). This complex is a SES if and only if for every prime ideal $\mathfrak{p} \triangleleft R$, the complex

$$0 \longrightarrow M_{1,\mathfrak{p}} \xrightarrow{f_\mathfrak{p}} M_{2,\mathfrak{p}} \xrightarrow{g_\mathfrak{p}} M_{3,\mathfrak{p}} \longrightarrow 0 \tag{4}$$

is a SES.

Proof. Functoriality of localization implies that if (3) is a complex then (4) is a complex for all primes $\mathfrak{p}$. Localization being an exact functor (Proposition 3.2.3) implies that if (3) is exact then (4) is exact for all primes $\mathfrak{p}$.

In order the prove the converse, we will take advantage of some very useful properties of localization:

- *If $N \subset M$ are R modules then $N_\mathfrak{p} \subset M_\mathfrak{p}$ and $(M/N)_\mathfrak{p} \cong M_\mathfrak{p}/N_\mathfrak{p}$.*
  Indeed, the sequence
  $$0 \longrightarrow N \longrightarrow M \longrightarrow M/N \longrightarrow 0$$
  is SES. Thus, also
  $$0 \longrightarrow N_\mathfrak{p} \longrightarrow M_\mathfrak{p} \longrightarrow (M/N)_\mathfrak{p} \longrightarrow 0$$
  is SES. This proves both statements.
- *Let $g\colon M \to M'$ be a homomorphism of R modules. Then*
  $$\ker(g_\mathfrak{p}) = \ker(g)_\mathfrak{p}.$$
  To show that we may assume w.l.o.g. that g is surjective and so
  $$0 \longrightarrow \ker(g) \longrightarrow M \xrightarrow{g} M' \longrightarrow 0$$
  is a SES. Thus, also
  $$0 \longrightarrow \ker(g)_\mathfrak{p} \longrightarrow M_\mathfrak{p} \xrightarrow{g_\mathfrak{p}} M'_\mathfrak{p} \longrightarrow 0$$
  is a SES and that proves our claim.
- *Let $g\colon M \to M'$ be a homomorphism of R modules. Then*
  $$\mathrm{Im}(g_\mathfrak{p}) = \mathrm{Im}(g)_\mathfrak{p}.$$
  This is straightforward: Let $S = R \setminus \mathfrak{p}$. Then, $\mathrm{Im}(g_\mathfrak{p}) = \{g(m)/s : m \in M, s \in S\} = g(M)_\mathfrak{p}$.

Returning to the proof, we assume that (4) is exact for every prime ideal $\mathfrak{p}$. Then $\ker(f_\mathfrak{p}) = \ker(f)_\mathfrak{p} = 0$ for all $\mathfrak{p}$ and, as being zero is a local property, $\ker(f) = 0$ and f is injective. Similarly, $M_{3,\mathfrak{p}} = \mathrm{Im}(g)_\mathfrak{p} = \mathrm{Im}(g_\mathfrak{p})$, implying that $(M_3/g(M_2))_\mathfrak{p} \cong M_{3,\mathfrak{p}}/g(M_2)_\mathfrak{p} = M_{3,\mathfrak{p}}/g_\mathfrak{p}(M_{2,\mathfrak{p}}) = 0$ for primes $\mathfrak{p}$. This implies that $M_3/g(M_2) = 0$ and so that g is surjective.

We are given $\mathrm{Im}(f) \subseteq \ker(g)$, and $(\ker(g)/\mathrm{Im}(f))_\mathfrak{p} \cong \ker(g)_\mathfrak{p}/\mathrm{Im}(f)_\mathfrak{p} \cong \ker(g_\mathfrak{p})/\mathrm{Im}(f_\mathfrak{p}) = 0$ for all $\mathfrak{p}$. Therefore, $\ker(g) = \mathrm{Im}(f)$. □

Of course, not every property of modules is a local property. For example, being free, or cyclic, is not a local property. See the exercises for an example.

4. Limits in a category

We are back to developing some concepts in category theory and, as usual, that means a lot of new terminology. We remark that we will be developing here the bare minimum – only that which is required for the course and its follow-up course. The whole subject can be done in greater generality and if you are interested look up books in category theory. We also remark that we are using “direct limit" and “injective limit" as synonyms. The same applies to “inverse limit" and “projective limit".

4.1. **Direct and inverse systmes.** Let I be a poset and $\mathbf{C}$ a category. We may view I as a category with objects being the elements of I and for $x, y \in I$,

$$\mathrm{Mor}(x,y) = \begin{cases} i_{xy} & x \leq y \\ \emptyset & else. \end{cases}$$

Composition is defined the only possible way: if $x \leq y \leq z$ then $i_{yz} \circ i_{xy} = i_{xz}$. To avoid confusion, we denote the category $\underline{I}$.

A **direct system** in $\mathbf{C}$ indexed by I is a *covariant* functor

$$\underline{I} \to \mathbf{C}.$$

This means exactly the following:

- For each $x \in I$, an object A_x of $\mathbf{C}$.
- For all $x \leq y$ in I, a morphism $f_{xy} \colon A_x \to A_y$ in $\mathbf{C}$ such that $f_{xx} = Id_{A_x}$.
- For all $x \leq y \leq z$ in I, $f_{yz} \circ f_{xy} = f_{xz}$.

An **inverse system** in $\mathbf{C}$ indexed by I is a *contravariant* functor

$$\underline{I} \to \mathbf{C}.$$

This means exactly the following:

- For each $x \in I$, an object A_x of $\mathbf{C}$.
- For all $x \leq y$ in I, a morphism $f_{xy} \colon A_y \to A_x$ in $\mathbf{C}$ such that $f_{xx} = Id_{A_x}$.
- For all $x \leq y \leq z$ in I, $f_{xy} \circ f_{yz} = f_{xz}$.

4.1.1. *Examples.*

(1) The constant system (either direct, or inverse): I is any poset, A an object of $\mathbf{C}$.

$$A_x = A, \ \forall x \in I; \qquad f_{xy} = Id_A, \ \forall x \leq y \in I.$$

(2) If I has the so-called **trivial**, or **discrete**, partial order – meaning, the only relations are $x \leq x$ for $x \in I$, then a direct (or inverse) system is just a collection of objects $\{A_x : x \in I\}$ in the category $\mathbf{C}$ with no maps between them except for Id_{A_x} for every x. Amusingly, this example is very important! It will define the notions of sums and products in a category.

(3) Another important example is the case where $I = \{1,2,3\}$ with the relation $1 \leq 2, 1 \leq 3$ (but 2 and 3 are in incomparable). In this case, a direct system indexed by I is a diagram in the category $\mathbf{C}$

$$\begin{array}{ccc} F_1 & \longrightarrow & F_2 \\ \downarrow & & \\ F_3 & & \end{array}$$

An inverse system indexed by I is

$$\begin{array}{ccc} & & F_2 \\ & & \downarrow \\ F_3 & \longrightarrow & F_1 \end{array}$$

(4) A poset I is called **directed** if for all $x, y \in I$ there is $z \in I$ such that $x \leq z$ and $y \leq z$. For example, take $I = \{1 < 2 < 3 < 4 \cdots\}$. A direct system indexed by I is

$$A_1 \to A_2 \to A_3 \to A_4 \cdots,$$

while an inverse system indexed by I is

$$A_1 \leftarrow A_2 \leftarrow A_3 \leftarrow A_4 \cdots$$

For another example, take $I = \{1,2,3,4,\dots\}$ but where now we declare that $m \leq n$ if $m|n$. Then a direct system indexed by I is a diagram

$$\begin{array}{ccccccc} A_1 & \longrightarrow & A_2 & \longrightarrow & A_4 & \longrightarrow & A_8 \\ & \searrow & & \searrow & & \searrow & \\ & & A_3 & \longrightarrow & A_6 & \longrightarrow & A_{12} \qquad \cdots \\ & \searrow & & \searrow & & \searrow & \\ & & A_5 & & A_9 & \longrightarrow & A_{18} \\ & \searrow & & \searrow & & & \\ & & A_7 & & A_{10} & \longrightarrow & A_{20} \\ & \searrow & & \searrow & & & \\ & & \vdots & & A_{14} & & \end{array}$$

Note that we don't indicate an arrow $A_2 \to A_8$, say. Such an arrow exists but it is necessarily the composition $A_2 \to A_4 \to A_8$ and so can inferred from the arrows that are already indicated.

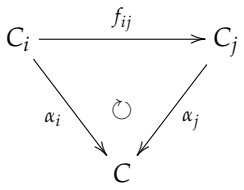
For an inverse system, reverse all the arrows.

We will denote a direct, or inverse, system by $(\{C_i\}, \{f_{ij}\})$ (where the C_i are objects of $\mathbf{C}$, one for every $i \in I$, and for $i \leq j$, f_{ij} are morphisms either from C_i to C_j, or conversely, depending if it is a direct or an inverse system).

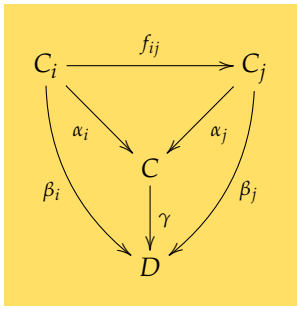
4.2. **Direct and inverse limits.** Let $(\{C_i\}, \{f_{ij}\})$ be a direct system in a category $\mathbf{C}$. A **direct limit** is an object C of $\mathbf{C}$ together with morphisms

$$\alpha_i \colon C_i \to C,$$

such that $\forall i \leq j$ we have commutative diagrams



and given an object D in $\mathbf{C}$ with morphisms $\beta_i \colon C_i \to D$, such that $\beta_j \circ f_{ij} = \beta_i$, there is a *unique* morphism $\gamma \colon C \to D$ such that the following diagrams commute $\forall i \leq j$:



meaning $\gamma \circ \alpha_i = \beta_i, \gamma \circ \alpha_j = \beta_j$.

We will denote the direct limit, if it exists (and that may depend on the system as well as the category) by

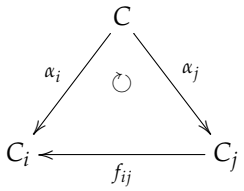
$$\varinjlim C_i$$

(or, if needed, $\varinjlim (\{C_i\}, \{f_{ij}\})$, or $\varinjlim_{i \in I} C_i$, and so on). If it exists it is unique up to unique isomorphism.

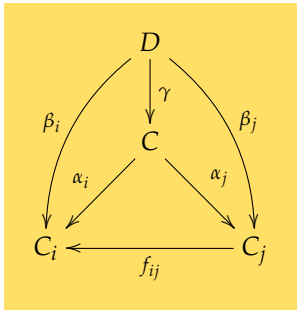
The definition of an inverse limit is quite similar. Let $(\{C_i\}, \{f_{ij}\})$ be an inverse system in a category $\mathbf{C}$. An **inverse limit** is an object C of $\mathbf{C}$ together with morphisms

$$\alpha_i \colon C \to C_i,$$

such that $\forall i \leq j$ we have commutative diagrams



and given an object D in $\mathbf{C}$ with morphisms $\beta_i\colon D \to C_i$, such that $f_{ij}\circ\beta_j = \beta_i$, there is a *unique* morphism $\gamma\colon D \to C$ such that the following diagrams commute $\forall i \le j$:



meaning $\alpha_i\circ\gamma = \beta_i, \alpha_j\circ\gamma = \beta_j$.

We will denote the inverse limit, if it exists (and that may depend on the system as well as the category) by

$$\varprojlim C_i$$

(or, if needed, $\varprojlim (\{C_i\},\{f_{ij}\})$, or $\varprojlim_{i\in I} C_i$, and so on). If it exists it is unique up to unique isomorphism.

In general, the issue of existence of limits can be rather subtle and even rather difficult. We provide here two very simple examples where limits do not exist, just so that we are aware that this is a potential issue.

Example 4.2.1. Let $\mathbf{C}$ have objects that are the finite sets $\{0,1,\cdots,n\}$, for all $n\in\mathbb{N}$. The morphisms are the natural inclusion maps. We have the direct system

$$\{0\}\to\{0,1\}\to\{0,1,2\}\to\{0,1,2,3\}\to\cdots$$

in $\mathbf{C}$. The direct limit does not exist. (We leave that as an exercise.)

For another example, let $\mathbf{C}$ be the category of all fields with ring homomorphisms between them. Let $I=\{2,3\}$ with the discrete order. Then $F_2=\mathbb{Z}/2\mathbb{Z}, F_3=\mathbb{Z}/3\mathbb{Z}$ is a direct system of fields indexed by I and a direct limit does not exists because there is no *field* F such that there are ring homomorphisms $F_2\to F$ and $F_3\to F$. We may consider the same system but now in the category of commutative rings. Then, a direct limit exists and it is simply $F_2\oplus F_3$ with the natural inclusion maps.

Theorem 4.2.2. *Let R be a ring. Inverse and direct limits exist in ${}_R\mathbf{Mod}$.*

Proof. Let $(\{M_i\},\{f_{ij}\})$ be a direct system in ${}_R\mathbf{Mod}$ indexed by I. Let

$$\oplus_{i\in I}M_i := \{(m_i)_{i\in I} : m_i\in M_i, m_i = 0 \text{ except for finitely many } i\}.$$

This is an R-module under the operations

$$(m_i)_i + (n_i)_i = (m_i+n_i)_i,\quad r(m_i)_i = (rm_i)_i.$$

In fact this module will turn out to be $\varinjlim M_i$ if I is discrete. Let

$$\lambda_i\colon M_i \to \oplus_{j\in I} M_j$$

be the inclusion in the i-th coordinate. Let $W\subseteq\oplus_{j\in I}M_j$ be the R-submodule generated by all expressions

$$\lambda_i(a) - \lambda_j(f_{ij}(a)),\quad i\le j\in I, a\in M_i.$$

Let

$$C = \oplus_{i\in I} M_i / W,$$

and $\alpha_i \colon M_i \to C$ the composition $M_i \xrightarrow{\lambda_i} \oplus_{j\in I} M_j \to C$. We claim that $(C, \{\alpha_i\})$ is a direct limit:

First, $\alpha_j \circ f_{ij} = \alpha_i$ because $\lambda_i(a) - \lambda_j(f_{ij}(a)) \in W, \forall a \in M_i$. Second, given $(D, \{\beta_i\})$ such that $\beta_j \circ f_{ij} = \beta_i$ define

$$\tilde{\gamma} \colon \oplus_{i\in I} M_i \to D, \quad \tilde{\gamma}((m_i)_i) = \sum_i \beta_i(m_i),$$

the sum being well-defined since almost all the m_i are zero. Then $\tilde{\gamma}(\lambda_i(a) - \lambda_j(f_{ij}(a)) = \beta_i(a) - \beta_j(f_{ij}(a)) = 0$. So $\tilde{\gamma}(W) = 0$ and we get a well-defined homomorphism of R-modules

$$\gamma \colon C \to D, \quad \gamma \circ \alpha_i = \beta_i$$

(because $\tilde{\gamma} \circ \lambda_i = \beta_i$.) Finally, note that since C is generated by the images of the α_i, γ is uniquely determined by the relation $\gamma \circ \alpha_i = \beta_i$.

Now, let $(\{M_i\}, \{f_{ij}\})$ be an inverse system in ${}_{\mathbf{R}}\mathbf{Mod}$ indexed by I. Let

$$\prod_{i\in I} M_i := \{(m_i)_{i\in I} : m_i \in M_i\}.$$

This is an R-module under the operations

$$(m_i)_i + (n_i)_i = (m_i + n_i)_i, \quad r(m_i)_i = (rm_i)_i.$$

In fact this module will turn out to be $\varprojlim M_i$ if I is discrete. Define $C \subseteq \prod_{i\in I} M_i$,

$$C = \{(m_i)_i : f_{ij}(m_j) = m_i, \forall i \le j\}.$$

C is an R-submodule of $\prod_{i\in I} M_i$. We have the projection maps

$$p_i \colon C \to M_i, \quad p_i((m_j)_j) := m_i,$$

which satisfy

$$f_{ij} \circ p_j = p_i.$$

Given $(D, \{q_i\})$, $q_i \colon D \to M_i, f_{ij} \circ q_j = q_i$, define

$$\gamma \colon D \to \prod_{i\in I} M_i, \quad \gamma(d) = (q_i(d))_i.$$

It is easy to check that in fact $\gamma \colon D \to C$ because $f_{ij}(q_j(d)) = q_i(d)$ and clearly $p_i \circ \gamma = q_i$. It remains to show γ is unique. If also $\delta \colon D \to C$, say $\delta(d) = (\delta_i(d))_i$, and δ satisfies $p_i \circ \delta = q_i$, then we get $\delta_i(d) = q_i(d)$ and so $\delta = \gamma$. □

4.2.1. *Examples.*

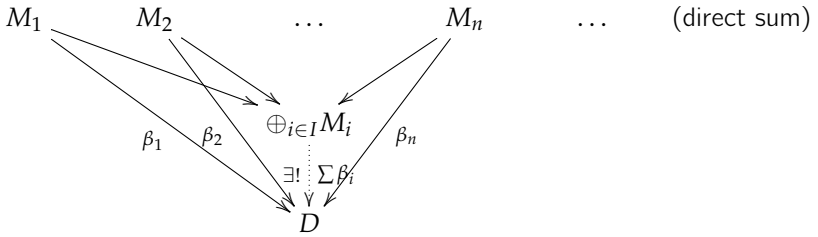
Example 4.2.3. (**Direct sum and direct product**) If I is discrete, we introduce the following notation and terminology for the category of R-modules:[6]

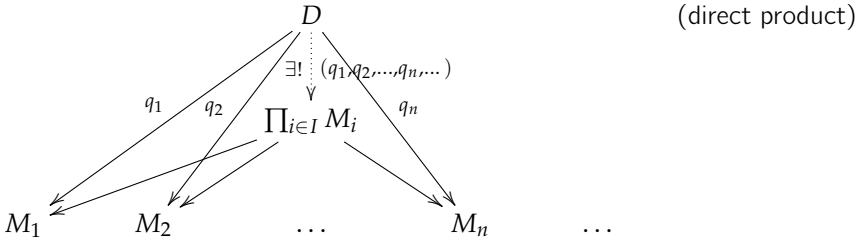
$$\varinjlim M_i = \oplus_{i\in I} M_i, \qquad \text{(direct sum)};$$

$$\varprojlim M_i = \prod_{i\in I} M_i, \qquad \text{(direct product)}.$$

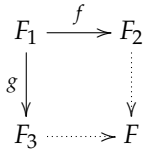
[6]The same definition and terminology can be used in any category; the notation might change. For example, the direct sum exists in the category of sets and is the disjoint union. In this case one uses the notation $\coprod$.

These are canonically isomorphic if I is finite, but not if I is infinite. Since these limits are so often used, we repeat their universal property:





Example 4.2.4. (Push-out) In a category $\mathbf{C}$ a **push-out** F is the following direct limit

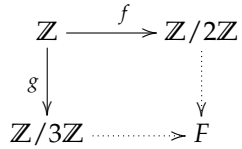


For ${}_{\mathbf{R}}\mathbf{Mod}$ we have

$$F \cong F_2 \oplus F_3/\{(f(a), -g(a)) : a \in F_1\}.$$

This is not precisely what we constructed and we leave it as an exercise to check the isomorphism.

To illustrate, let us compute the pushout of



where the maps f, g are the obvious maps sending an integer to its congruence class. The push-out is the zero ring as it is isomorphic to

$$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}/\{(a,-a) : a \in \mathbb{Z}\} \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}/\langle(1,-1)\rangle \cong \mathbb{Z}/6\mathbb{Z}/\langle 5\rangle,$$

since under the Chinese remainder theorem, the element $5 \pmod 6$ is sent to $(5,5) = (1,-1)$ in $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$. Since 5 generates $\mathbb{Z}/6\mathbb{Z}$, that quotient is 0. This is a simple example, but you may want to revisit it once we have studied the spectra of affine rings to understand the geometry behind (the intersection of distinct points in $\mathrm{Spec}(\mathbb{Z})$ is empty) and see how to generalize the example to a ring R with two quotients $R/\mathfrak{p}, R/\mathfrak{q}$ where $\mathfrak{p}$ and $\mathfrak{q}$ are, distinct prime ideals, say.

Example 4.2.5. (Pull-back) In a category $\mathbf{C}$ the **pull-back** F is the following inverse limit

$$\begin{array}{ccc} F & \dashrightarrow & F_2 \\ \vdots & & \downarrow f \\ F_3 & \xrightarrow[g]{} & F_1 \end{array}$$

We say that the square is **cartesian**. In particular, one finds the following explicit descriptions:

- In **Sets**, $F = \{(a,b) \in F_2 \times F_3 : f(a) = g(b)\}$.
- In **Top**, F is the same, endowed with the subspace topology.
- In ${}_{\mathbf{R}}\mathbf{Mod}$, F is the same as in **Sets**, and it is a submodule of $F_2 \times F_3$.
- In **Gps**, F is the same as in **Sets**, and it is a subgroup of $F_2 \times F_3$.
- We shall see that in the category of affine schemes the pull-back (also called fibre product) behaves differently.

Example 4.2.6. (**limits over a directed set**) Consider the category ${}_{\mathbf{R}}\mathbf{Mod}$. If I is a directed index set (meaning, $\forall x,y \in I, \exists z \in I, x \leq z, y \leq z$) then $\varinjlim M_i$ has another description which is very convenient in applications. Consider expressions

$$(i,\alpha), \qquad i \in I, \alpha \in M_i$$

and define a relation by

$$(i,\alpha) \sim (j,\beta)$$

if $\exists k$ such that $i \leq k, j \leq k$ and

$$f_{ik}(\alpha) = f_{jk}(\beta).$$

So, in particular $(i,\alpha) \sim (k, f_{ik}(\alpha)) = (k, f_{jk}(\beta)) \sim (j,\beta)$. This is an equivalence relation. The equivalence classes are an R-module under the definitions

$$r \cdot (i,\alpha) = (i, r\alpha), \quad r \in R,$$

and

$$(i,\alpha) + (j,\beta) = (k, f_{ik}(\alpha) + f_{jk}(\beta)),$$

for any k such that $i \leq k, j \leq k$.

Example 4.2.7. $I = \{1,2,3,\dots\}$ with order relation of division, $m \leq n$ if $m|n$. Consider the direct system

$$\{(M_m, f_{mn}), \quad M_m = \mathbb{Z}, \quad f_{mn}(a) = \tfrac{m}{n} \cdot a\}.$$

It is not hard to show that this direct system is isomorphic to

$$\{(M_m, f_{mn}), \quad M_m = \tfrac{1}{m}\mathbb{Z}, \quad f_{mn}(a) = a\}.$$

Using that I is a directed poset, we can use the alternative description of the direct limit and conclude that

$$\varinjlim \tfrac{1}{m}\mathbb{Z} \cong \mathbb{Q}.$$

Example 4.2.8. (**Ring of germs of functions**) Let V be a real manifold (or a complex manifold, or just a topological space) and let $v_0 \in V$. Let $\mathbf{C}$ be the category of commutative rings. Let I be the collection of open sets U in V that contain v_0. It is considered as directed index set where

$$U_1 \leq U_2 \quad \text{if} \quad U_2 \subset U_1.$$

For $U \in I$, we let

$$R(U) = \text{differential functions } U \to \mathbb{R}$$

(resp., holomorphic functions; resp., continuous functions) and for $U_1 \leq U_2$ we have the canonical ring homomorphism

$$R(U_1) \to R(U_2), \quad f \mapsto f|_{U_2}.$$

Then $\varinjlim R(U)$, the limit taken over all open sets U containing v_0, is called the **local ring of germs of differential functions at** v. Heuristically we think about it as the ring of differentiable functions that are defined on some open, unspecified, set containing v_0. It is indeed a local ring; its maximal ideal are the functions vanishing at v_0.

Sometimes it can made even more concrete. For example, if $V = \mathbb{C}$ and we are considering germs of holomorphic functions at 0, then the ring of germs can be identified with

$$\{\sum_{n=0}^{\infty} a_n z^n : a_n \in \mathbb{C}, \exists \epsilon > 0, \sum_{n=0}^{\infty} |a_n| \epsilon^n < \infty\}.$$

Example 4.2.9. (I-**adic completion**) Let R be a commutative ring and take the index set $\{1 < 2 < 3 < \dots\}$. Let $I \neq R$ be an ideal of R (so *not* the index set this time). We have an inverse system index by $\{1 < 2 < 3 < \dots\}$

$$R/I \longleftarrow R/I^2 \longleftarrow R/I^3 \longleftarrow \cdots$$

In this case we introduce the following notation and terminology

$$\varprojlim R/I^n =: R^{\wedge I} = \text{ the } I\text{-adic completion of } R \text{ at } I.$$

It is the ring of "consistent vectors"

$$R^{\wedge I} = \{(m_a)_{a=1}^{\infty} : m_a \in R/I^a, m_{a+1} \equiv m_a \pmod{I^a}\}.$$

We know $R^{\wedge I}$ is an R-module, but it is also a ring:

$$(m_i)_i \cdot (n_i)_i = (m_i n_i)_i.$$

Further, there is a natural ring homomorphism

$$R \to R^{\wedge I}, \qquad r \mapsto (r, r, r, \dots).$$

The kernel of this map is $\cap_{a=1}^{\infty} I^a$, which is an ideal of R that need not be zero in general. However, we have the following important case (the notion of a noetherian ring will be defined later in § 7.1, but, for example, every quotient of a polynomial ring in several variables over a field is noetherian).

Theorem 4.2.10. (Krull's intersection theorem) *Let R be a noetherian ring, $I \lhd R$. Then the kernel of the map $R \to R^{\wedge I}$ are the elements $x \in R$ such that exists $r \in I$ with $(1-r)x = 0$. In fact, there exists $r \in I$ such that $(1-r)(\cap_{a=1}^{\infty} I^a) = \{0\}$. In particular, if R is an integral domain, or a local ring, and $I \neq R$ then $R \hookrightarrow R^{\wedge I}$.*

The device of completion, in particular when $R \hookrightarrow R^I$, allows one to do analysis in an algebraic setting. We illustrate this in two settings:

- **Power series**. Let k be a field and $R = k[x]$. Let $I = \langle x \rangle$ then

$$k[x]^{\wedge I} \cong k[[x]].$$

Briefly, $\sum_{i=0}^{\infty} a_i x^i$ defines a consistent vector $(\sum_{i=0}^{n-1} a_i x^i)_n \in k[x]^{\wedge I}$, and conversely (use that any element in R/I^n has a unique representative $\sum_{i=0}^{n-1} a_i x^i$). This is quite similar to ring of germs of analytic functions we constructed for the local ring of $\mathbb{C}$ at 0, only that convergence does not play a factor algebraically (and, in fact, would not even make sense for a general field k).

- p-**adic numbers**. Let $R = \mathbb{Z}, I = (p)$, p a prime number. Then

$$R^{\wedge I} =: \mathbb{Z}_p = \text{ the } p\text{-adic numbers},$$

is a ring that is central in number theory.[7] We have the following remarkable facts, all at the level of an exercise,

$$\mathbb{Z} \subset \mathbb{Z}_p \underset{\text{closed}}{\subset} \prod_{n=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z}.$$

Each ring $\mathbb{Z}/p^n\mathbb{Z}$ is given here the discrete topology and the product is then a compact Hausdorff space by Tychonoff's theorem, thus $\mathbb{Z}_p$ is a compact Hausdorff topological ring that contains $\mathbb{Z}$. Further, every element of $\mathbb{Z}_p$ can be represented by

$$\sum_{i=0}^{\infty} a_i p^i, \quad a_i \in \{0, 1, \ldots, p-1\},$$

and that is true both as a formal sum where we identify such sums (as in the case of power series) with $\mathbb{Z}^{\wedge I}$ by

$$\sum_{i=0}^{\infty} a_i p^i \mapsto (\sum_{i=0}^{n-1} a_i p^i)_n \in \mathbb{Z}^{\wedge I},$$

and both as a limit $\varinjlim_n \sum_{i=0}^{n-1} a_i p^i$ taken in the topological ring $\mathbb{Z}^{\wedge I}$. As a consequence we see that $\mathbb{Z}$ is dense in $\mathbb{Z}_p$. In fact, more is true. One can define a metric on $\mathbb{Z}$ by[8]

$$d(a, b) = p^{-n}, \quad p^n \| (a - b).$$

It then turns out that $\mathbb{Z}_p$ is the metric completion of $\mathbb{Z}$. The extension of the metric to $\mathbb{Z}_p$ is provided by

$$d\left(\sum_{i=0}^{\infty} a_i p^i, \sum_{i=0}^{\infty} b_i p^i\right) = p^{-t}, \quad t = \min\{i : a_i \neq b_i\}.$$

Example 4.2.11. (**Kernels and cokernels**) Let $f\colon M \to N$ be a homomorphism of R-modules. Then $\mathrm{Ker}(f)$ is the pull-back

$$\begin{array}{ccc} \mathrm{Ker}(f) & \longrightarrow & \{0\} \\ \downarrow & & \downarrow \\ M & \xrightarrow{f} & N \end{array}$$

This allows defining $\mathrm{Ker}(f)$ without using elements in settings where we do have a zero object and final inverse limits. That is, it avoids defining the kernel as the set of element $\{m \in M : f(m) = 0\}$ and so can

[7] Do not confuse it with the localization $\mathbb{Z}_{(p)}$, which is localizing that multiplicative set $\mathbb{Z} - (p)$. Also, some authors use $\mathbb{Z}_p$ to denote $\mathbb{Z}/p\mathbb{Z}$, a finite field with p-elements. Again, totally different.

[8] The expression $p^n \| (a - b)$ means that p^n divides $a - b$ but p^{n+1} does not.

be applied to categories where morphisms aren't necessarily functions. This leads to the notion of abelian category.

Similarly, Coker(f) $= N/f(M)$ can be defined as the push-out

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ \downarrow & & \downarrow \\ 0 & \longrightarrow & \mathrm{Coker(f)} \end{array}$$

Example 4.2.12. (**Amalgamated product**) Let

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\downarrow & & \\ C & & \end{array}$$

be a diagram of (not necessarily) commutative groups and group homomorphisms. The push-out of this diagram exists in **Gps**. It is denoted

$$B \underset{A}{*} C,$$

and called the **amalgamated product** of B and C over A. We remark that if $A = \{1\}$ the push-out is the same as the direct limit of the discrete system $(\{B, C\})$. In this case we write $B * C$ and call it the **free product**.

$B *_A C$ can be constructed as the group whose elements are equivalence classes of finite expressions ("words")

$$x_1 x_2 \dots x_n, \quad x_i \in B \coprod C,$$

subject to the relations:

(1) If x_i, x_{i+1} belong to the same group $x_1 x_2 \dots x_n = x_1 x_2 \dots x_{i-1}(x_i x_{i+1}) x_{i+2} \dots x_n$.
(2) In particular, if $x_i = 1$, $x_1 x_2 \dots x_n = x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_n$.
(3) If $x_i \in B, x_{i+1} \in C$ then for every $a \in A$, we have the relation $x_1 x_2 \dots (x_i f(a)) x_{i+1} \dots x_n = x_1 x_2 \dots x_i (g(a) x_{i+1}) \dots x_n$. And similarly if $x_i \in C, x_{i+1} \in B$.

Multiplication is defined by concatenation of words.

Here are some interesting examples of this consruction:

- If $A = \{1\}$ we get the free product $B * C$. For example, $\mathbb{Z} * \mathbb{Z}$ is the free group on 2 generators.
- One can prove that $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z} \cong \mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm I_2\}$, which is a group with a very rich structure!
- Let $Z = X \cup Y$, where X, Y, Z are all connected, locally path connected, Hausdorff topological spaces and X, Y are open in Z. We also assume that $X \cap Y$ is connected and locally path connected. Choose a base point $t \in X \cap Y$. We then have

  Theorem 4.2.13. (Seifert & Van-Kampen)

  $$\pi_1(Z, t) \cong \pi_1(X, t) \underset{\pi_1(X \cap Y, t)}{*} \pi_1(Y, t).$$

  Using this theorem, it is easy to deduce that the fundamental group of two circles touching at a point is a free group on 2 generators and the fundamental group of a two-dimensional sphere is trivial.

4.3. **Limits and adjoint functors.** We prove study here the interaction of functors with limits. For brevity, we only consider covariant functors, but similar statements exists for contravariant functors.

Theorem 4.3.1. *Let (F,G) be an adjoint pair of covariant functors, $F\colon \mathbf{C}\to\mathbf{D}, G\colon \mathbf{D}\to\mathbf{C}$, then F preserves all direct limits and G preserves all projective limits.*

In order the prove the theorem we need to go a bit further in our study of adjoint functors. In general, for an object $C\in\mathbf{C}$, GFC is not isomorphic to C. For instance, we have the adjoint pair (ab, inc) where for a group G, $inc(ab(G)) = G^{ab}$ which is not isomorphic to G unless G happens to be abelian. Nonetheless, we show now that there is a canonical connection between C and GFC:

To begin with, recall that the adjoint property means that we have natural isomorphisms

$$\mathrm{Mor}_{\mathbf{D}}(FA,B)\xrightarrow[\sim]{\varphi_{A,B}}\mathrm{Mor}_{\mathbf{C}}(A,GB)\ .$$

Letting $B=FA$, we find natural isomorphisms

$$\mathrm{Mor}_{\mathbf{D}}(FA,FA)\xrightarrow[\sim]{\varphi_{A,FA}}\mathrm{Mor}_{\mathbf{C}}(A,GFA)\ .$$

In particular, we get

$$\eta_A := \varphi_{A,FA}(1_{FA})\in\mathrm{Mor}_{\mathbf{C}}(A,GFA).$$

In fact, η gives us a natural transformation of functors $\mathbf{C}\to\mathbf{C}$[9]

$$\eta : \mathbb{1}_{\mathbf{C}}\to GF;$$

for every object A of $\mathbf{C}$ we have a morphism $\eta_A\colon A\to GFA$ and this morphism is natural as the following diagram commutes

$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & GFA \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle GFf} \\ A' & \xrightarrow{\eta_{A'}} & GFA' \end{array}$$

(The commutativity follows from $\varphi_{A,B}$ being natural in both A and B, a property we apply for both $f\colon A\to A'$ and $GFf\colon GFA\to GFA'$.) The natural transformation η is called the **unit** of the adjoint pair.

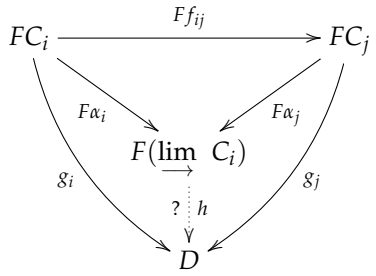
Proof. (Theorem) We sketch the proof for F, leaving completing the proof and providing a proof for G as an exercise (that requires defining a co-unit, etc.). Let I be a poset and (C_i, f_{ij}) a direct system in $\mathbf{C}$ indexed by I. Applying F we get a direct system (FC_i, Ff_{ij}) in $\mathbf{D}$. We have diagrams

$$\begin{array}{ccccc} FC_i & & \xrightarrow{\quad Ff_{ij}\quad} & & FC_j \\ & {\scriptstyle F\alpha_i}\searrow & \circlearrowleft & \swarrow{\scriptstyle F\alpha_j} & \\ & & F(\varinjlim\ C_i) & & \end{array}$$

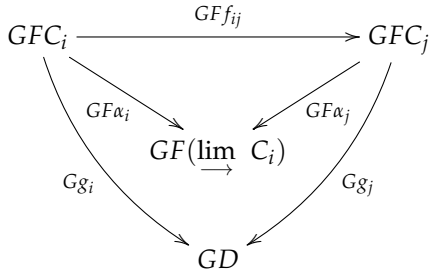
[9]One can also take $A=GB$ and define a **co-unit** $\epsilon_B : \mathrm{Mor}_{\mathbf{D}}(FGB,B)$ and use it to prove the statement about projective limits.

Suppose we have an object D of $\mathbf{D}$ and commutative diagrams

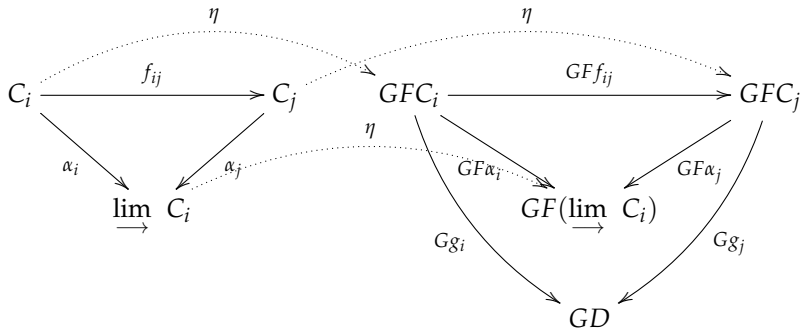
(5)



where we are looking to prove the existence of a unique dotted arrow h making those diagrams commutative. Apply G to get

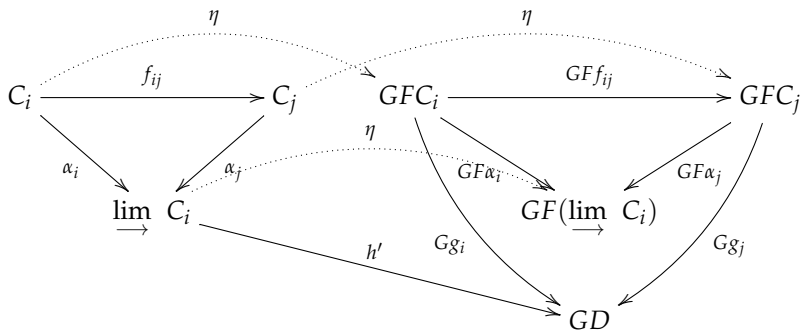


The problem at this point is that GF is not $\mathbb{1}_{\mathbf{C}}$, so we don't get back to the original direct limit. But we *can* relate GF to $\mathbb{1}_{\mathbf{C}}$ using the unit η of the adjoint pair. So we have



Here, we didn't include the subscripts for the maps η and we made some horizontal arrows dotted – they are well-defined, the dots are just to make the diagram easier to read. Using the universal property of the direct limit, and the functoriality of η, we get a unique morphism $h' : \varinjlim C_i \to GD$ that makes the subdiagram with maps $Gg_i \circ \eta, Gg_j \circ \eta, \alpha_i, \alpha_j$ commutative.

(6)



Under the isomorphism $\mathbf{Mor}_{\mathbf{C}}(\varinjlim C_i, GD) \cong \mathbf{Mor}_{\mathbf{D}}(F\varinjlim C_i, D)$, h' corresponds to a morphism $h \in \mathbf{Mor}_{\mathbf{D}}(F\varinjlim C_i, D)$. but it is not clear that the diagram (5) commutes.

Firstly, the following diagram, which is commutative by definition of adjoint,

$$\begin{array}{ccc} \mathbf{Mor}_{\mathbf{D}}(FC_i, D) & \xrightarrow{\cong} & \mathbf{Mor}_{\mathbf{C}}(C_i, GD) \\ \downarrow{\scriptstyle \circ F\alpha_i} & & \downarrow{\scriptstyle \circ \alpha_i} \\ \mathbf{Mor}_{\mathbf{D}}(F\varinjlim C_i, D) & \xrightarrow{\cong} & \mathbf{Mor}_{\mathbf{C}}(\varinjlim C_i, D) \end{array}$$

allows us to conclude at the level of the top row that $h \circ F\alpha_i$ corresponds to $h' \circ \alpha_i = h' \circ (\alpha_j \circ f_{ij})$ that corresponds to $h \circ F(\alpha_j \circ f_{ij}) = h \circ F\alpha_j \circ Ff_{ij}$. So, $h \circ F\alpha_i = h \circ F\alpha_j \circ Ff_{ij}$. Second, similarly, $h \circ F\alpha_i$ corresponds to $h' \circ \alpha_i = Gg_i \circ \eta$. Consider the diagram

$$\begin{array}{ccc} \mathbf{Mor}_{\mathbf{C}}(C_i, GFC_i) & \xrightarrow{Gg_i\circ} & \mathbf{Mor}_{\mathbf{C}}(C_i, GD) \\ \downarrow{\scriptstyle \cong} & & \downarrow{\scriptstyle \cong} \\ \mathbf{Mor}_{\mathbf{D}}(FC_i, FC_i) & \xrightarrow{g_i\circ} & \mathbf{Mor}_{\mathbf{D}}(FC_i, D) \end{array}$$

$Gg_i \circ \eta$ in the upper right corner, comes from η in the upper left corner that corresponds to 1_{FC} in the lower left corner, which corresponds to g_i in the lower right corner. Namely, $Gg_i \circ \eta$ also corresponds to $g_i \in \mathbf{Mor}_{\mathbf{D}}(FC_i, D)$ and so must be $h \circ F\alpha_i$.

The proof of unicity of h is proved using these same diagrams and consists in showing that under the isomorphism $\mathbf{Mor}(\varinjlim C_i, GD) \cong \mathbf{Mor}(F\varinjlim C_i, D)$, any other map $h_1 \in \mathbf{Mor}(F\varinjlim C_i, D)$ that makes (5) commutative corresponds to $h_1' \in \mathbf{Mor}(\varinjlim C_i, GD)$ that satisfies the appropriate commutative diagram, forcing it to be h'. □

Corollary 4.3.2. *Let $B \in {}_S\mathbf{Mod}_R$. The functor $B \otimes_R (-)$ from R-modules to S-modules commutes with arbitrary direct limits. In particular $B \otimes_R (\oplus_i A_i) \cong \oplus_i B \otimes_R A_i$.*

The functor $\mathrm{Hom}_S(B, -)$ from S-modules to R-modules commutes with arbitrary inverse limits. In particular, $\mathrm{Hom}_S(B, \prod_i A_i) \cong \prod_i \mathrm{Hom}_S(B, A_i)$.[10]

An interesting example of the corollary is to use it to analyze $B \otimes_{\mathbb{Z}} \mathbb{Q}$ and $\mathrm{Hom}(B, \mathbb{Z}_p)$, for an abelian group B, making use of the expressions $\mathbb{Q} = \varinjlim \frac{1}{n!}\mathbb{Z}$ and $\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}$. We leave that to the reader.

[10] The statements are also quite clear from the definitions, so using the adjoint property in this case is a bit of an overkill.

Part 2. SPECTRA, INTEGRABILITY AND NOETHERIANITY

5. The spectrum of a ring

Let R be a commutative ring. Alexander Grothendieck associated to R a special kind of topological space denoted $\mathrm{Spec}(R)$, the spectrum of R. It has much more structure than merely a topology. It is a so-called locally ringed space. This definition was the basis for a revolution in algebraic geometry. In a blink of an eye algebraic geometry was extended from varieties that were associated to very special rings of the form $k[x_1,\dots,x_n]/\langle f_1,\dots,f_n\rangle$ for some polynomials $f_1,\dots,f_n$ with coefficients in a field k, to a vast universe where any commutative ring is allowed. Besides a mere generalization this allowed the resolution of important foundational issues in algebraic geometry; for a long time the development of algebraic geometry was spearheaded by the "great Italian geometers", but a crisis in its foundations was brewing. Grothendieck's work, the bulk of which was done in the 1950's and 60's at the IHES with assistance of many of the brightest minds of that era, allowed to re-write classical algebraic geometry and put it on solid foundations. From a different perspective, it allowed a new kind of geometry that was essential to the development of arithmetic geometry.[11] Grothendieck's work fully realized the beautiful saying of Sophie Germain (1776-1831):

"L'algèbre n'est qu'une géeométrie écrite; la géométrie n'est qu'une algèbre figurée".

5.1. $\mathrm{Spec}(R)$ as a set.

Let R be a commutative ring. We define the set

$$\mathrm{Spec}(R) = \{[\mathfrak{p}] : \mathfrak{p} \text{ a prime ideal of } R\}.$$

Namely, the points in $\mathrm{Spec}(R)$ are the prime ideals of R. We use the brackets so as to distinguish between the point $[\mathfrak{p}]$ in the set $\mathrm{Spec}(R)$ and the actual prime ideal $\mathfrak{p}$.

Example 5.1.1. Here are a few simple examples:

(1) If R is a field, $\mathrm{Spec}(R) = \{[0]\}$. It is a singleton corresponding to the ideal $\{0\}$ of R.
(2) Likewise, if R is a field and t is a variable, then $\mathrm{Spec}(R[t]/(t^2))$ is a singleton corresponding to the unique prime ideal of this ring, the ideal (t). One often writes this ring as $R[\epsilon]$ where it is understood that $\epsilon^2 = 0$. It is called the **ring of dual numbers**.
(3) $\mathrm{Spec}(\mathbb{Z}) = \{[0], [2\mathbb{Z}], [3\mathbb{Z}], [5\mathbb{Z}], \dots\}$.
(4) $\mathrm{Spec}(\mathbb{C}[x]) = \{[0]\} \cup \{[(x-\alpha)] : \alpha \in \mathbb{C}\}$. Let us explain that:
    Suppose that I is a prime ideal of $\mathbb{C}[x]$ that is not 0. Then I contains some polynomial $f(x)$ that is not constant. As I is prime, it contains also some prime factor of f and so we may assume that f is irreducible. But such f must be a degree 1 polynomial. So $I \supseteq (x-\alpha)$ for some α. As the ideal $(x-\alpha)$ is maximal (the quotient $\mathbb{C}[x]/(x-\alpha) \cong \mathbb{C}$), it follows that $I = (x-\alpha)$.
(5) $\mathrm{Spec}(\mathbb{R}[x])$ is a bit more complicated. Besides points of the form $[(x-\alpha)], \alpha \in \mathbb{R}$ and $[0]$, it also contains points of the form $[(x^2+bx+c)]$ with $b^2-4c<0$. These are all the points of $\mathrm{Spec}(\mathbb{R}[x])$.

The following lemma is clear.

Lemma 5.1.2. *A homomorphism of rings $f: R \to S$ induces a function*

$$f^*: \mathrm{Spec}(S) \to \mathrm{Spec}(R), \qquad [\mathfrak{p}] \mapsto [f^{-1}(\mathfrak{p})].$$

[11] See, for example, *Intuition and Rigor and Enriques's Quest* by David Mumford, Notices AMS 2011 and *How Grothendieck Simplified Algebraic Geometry* by Colin McLarty, Notices AMS 2016.

Corollary 5.1.3. *$R \mapsto \mathrm{Spec}(R)$ is a contravariant functor* **CommRings** $\to$ **Sets**.

Example 5.1.4.
(1) The homomorphism $\mathbb{Z} \to \mathbb{F}_p$, embeds the point $\mathrm{Spec}(\mathbb{F}_p)$ as the point $[p\mathbb{Z}] \in \mathrm{Spec}(\mathbb{Z})$.
(2) The homomorphism $\mathbb{R}[x] \to \mathbb{C}[x]$ induces a function $\mathrm{Spec}(\mathbb{C}[x]) \to \mathrm{Spec}(\mathbb{R}[x])$ that takes the point $[(x-\alpha)]$ to the point $[(x-\alpha)]$ if $\alpha \in \mathbb{R}$ and to the point $[(x-\alpha)(x-\bar{\alpha})]$ if $\alpha \notin \mathbb{R}$.
(3) Let $h \in R$ and $S = \{1, h, h^2, \dots\}$ a multiplicative set. We will denote the localization $R[S^{-1}]$ by $R[h^{-1}]$. It is in fact isomorphic to $R[x]/(xh-1)$. By Theorem 3.3.1, the ring homomorphism

$$\ell \colon R \to R[h^{-1}], \quad \ell(r) = \frac{r}{1},$$

gives an injective map $\mathrm{Spec}(R[h^{-1}]) \to \mathrm{Spec}(R)$ whose image is the set of prime ideals of R that do not contain h.

5.2. $\mathrm{Spec}(R)$ **as a topological space.** Before defining the topology on $\mathrm{Spec}(R)$, we consider the notion of a radical of an ideal. Let $\mathfrak{a} \lhd R$ be an ideal. Its **radical** $\sqrt{\mathfrak{a}}$ is

$$\sqrt{\mathfrak{a}} = \{r \in R : \exists n \geq 1, r^n \in \mathfrak{a}\}.$$

It is indeed an ideal. If $r \in \sqrt{a}$ and $r^n \in \mathfrak{a}$ then for every $s \in R$ also $(sr)^n = s^n r^n \in \mathfrak{a}$ and so $R\sqrt{\mathfrak{a}} \subseteq \sqrt{\mathfrak{a}}$. If $r_i \in \sqrt{\mathfrak{a}}$ and say $r_i^{n_i} \in \mathfrak{a}$ then $(r_1 + r_n)^{n_1+n_2}$ is, by the binomial formula, a sum of terms each of which is either a multiple of $r_1^{n_1}$ or of $r_2^{n_2}$, thus in $\mathfrak{a}$. Therefore, $\sqrt{a}$ is an ideal containing $\mathfrak{a}$.

An ideal $\mathfrak{a}$ is called a **radical ideal** if $\mathfrak{a} = \sqrt{\mathfrak{a}}$. For example, if $\mathfrak{a}$ is a prime ideal then it is a radical ideal. Also, for every ideal $\mathfrak{a}$, $\sqrt{\mathfrak{a}}$ is a radical ideal.

Now let $\mathfrak{a}$ be an ideal of R. Define a subset of $\mathrm{Spec}(R)$:

$$V(\mathfrak{a}) = \{[\mathfrak{p}] \in \mathrm{Spec}(R) : \mathfrak{p} \supseteq \mathfrak{a}\}.$$

We define a topology on $\mathrm{Spec}(R)$ whose closed sets are the $V(\mathfrak{a})$. Note that $V(\{0\}) = \mathrm{Spec}(R)$ and $V(R) = \emptyset$. To show this is a topology, we must prove assertions (1) & (3) of following proposition.

Proposition 5.2.1. *The following holds:*
(1) $V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a}\mathfrak{b})$.
(2) $V(\mathfrak{a}) = V(\mathfrak{b}) \Leftrightarrow \sqrt{\mathfrak{a}} = \sqrt{\mathfrak{b}}$.
(3) $\cap_i V(\mathfrak{a}_i) = V(\sum_i \mathfrak{a}_i)$, *where* $\sum_i \mathfrak{a}_i$ *is the minimal ideal of* R *containing all the ideals* $\mathfrak{a}_i$.

Proof. We begin with the first claim. Any ideal containing $\mathfrak{a}$ (resp. $\mathfrak{b}$) contains $\mathfrak{a}\mathfrak{b}$, so the r.h.s contains the l.h.s. Let $\mathfrak{p} \supseteq \mathfrak{a}\mathfrak{b}$ be a prime ideal and suppose that $\mathfrak{p}$ does not contain $\mathfrak{b}$. There is thus $b \in \mathfrak{b}$ such that $b \notin \mathfrak{p}$. For every $a \in \mathfrak{a}$ the element $ab \in \mathfrak{a}\mathfrak{b}$ and so in $\mathfrak{p}$. As $\mathfrak{p}$ is prime and $b \notin \mathfrak{p}$ it must be that $a \in \mathfrak{p}$. It follows that $[\mathfrak{p}] \in V(\mathfrak{a})$.

For the second claim, we first note that if $\mathfrak{p} \supseteq \mathfrak{a}$ is a prime ideal then $\mathfrak{p} \supseteq \sqrt{a}$. Thus, $V(\mathfrak{a}) = V(\sqrt{\mathfrak{a}})$. So, it is enough to prove that the set of prime ideals $\mathfrak{p}$ that contain a given radical ideal determines it. In fact, we have the following lemma.

Lemma 5.2.2. *Let* $\mathfrak{a}$ *be an ideal of* R *then*

$$\sqrt{\mathfrak{a}} = \bigcap_{\mathfrak{p} \supseteq \mathfrak{a}} \mathfrak{p},$$

the intersection being over prime ideals containing $\mathfrak{a}$.

Proof. One inclusion is clear. If $w^n \in \mathfrak{a}$ and $\mathfrak{p} \supseteq \mathfrak{a}$ is a prime ideal then $w^n \in \mathfrak{p}$ and so $w \in \mathfrak{p}$. That gives $\sqrt{\mathfrak{a}} \subseteq \cap_{\mathfrak{p} \supseteq \mathfrak{a}} \mathfrak{p}$.

Now, take an element $f \in R$ such that $f \notin \sqrt{a}$. That is, for all $n > 0$, $f^n \notin \mathfrak{a}$. Let Σ be the set of ideals $\mathfrak{b} \supseteq \mathfrak{a}$ of R with the property that $f^n \notin \mathfrak{b}$ for all $n > 0$. This is a non-empty set as $\mathfrak{a} \in \Sigma$; it is partially ordered under inclusion and every chain of ideals $\{\mathfrak{b}_\alpha\}_{\alpha \in I}$ has an upper bound $\cup_{\alpha \in I} \mathfrak{b}_\alpha$ that belongs to Σ. By Zorn's Lemma, Σ has a maximal element $\mathfrak{p}$. We claim that $\mathfrak{p}$ is a prime ideal and being in Σ, $\mathfrak{p} \supseteq \mathfrak{a}$ and $f \notin \mathfrak{p}$. Thus, we are done.

Suppose that $\mathfrak{p}$ is not a prime ideal. Then, there are $x, y \in R$ such that

$$xy \in \mathfrak{p}, \quad x \notin \mathfrak{p}, y \notin \mathfrak{p}.$$

Therefore, we have a strict inclusion of ideals $(x) + \mathfrak{p} \supsetneq \mathfrak{p}$, $(y) + \mathfrak{p} \supsetneq \mathfrak{p}$. Thus, for some n, m positive integers we have $f^n \in (x) + \mathfrak{p}, f^m \in (y) + \mathfrak{p}$, But then, $f^{m+n} \in ((x) + \mathfrak{p})((y) + \mathfrak{p}) \subseteq \mathfrak{p}$. Contradiction. □

The last claim of the proposition is rather easy. If $\mathfrak{p}$ contains each $\mathfrak{a}_i$ it contains the minimal ideal containing all of them, and vice-versa. □

Note an interesting corollary of the lemma.

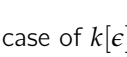
Corollary 5.2.3. *The* **nilradical** *of R, that is, the ideal $\sqrt{0}$, is the collection of all nilpotent elements of R and is equal to the intersection of all prime ideals of R:*

$$\sqrt{0} = \cap_{\mathfrak{p} \text{ prime}} \mathfrak{p} = \{r \in R : \exists n > 0, r^n = 0\}.$$

Returning to our main topic, we have established the existence of a topology on $\mathrm{Spec}(R)$ whose closed sets are the sets $V(\mathfrak{a})$.

Example 5.2.4. It is not hard to show that if $\mathfrak{p}$ is a prime ideal then $V([\mathfrak{p}])$ is the closure of the one point set $\{[\mathfrak{p}]\}$. Let's consider a few simple examples:

- If k is a field then $\mathrm{Spec}(k), \mathrm{Spec}(k[\epsilon]), \mathrm{Spec}(k[x,y,z]/(x^2, y^3, z^{11})$ are just one point spaces.

  where $\mathfrak{p} = (0)$ in the case of k, (ϵ) in the case of $k[\epsilon]$ and (x,y,z) for the last ring.

- Let p be a prime and consider $\mathrm{Spec}(\mathbb{Z}_{(p)})$, where $\mathbb{Z}_{(p)}$ is the localization of $\mathbb{Z}$ in the multiplicative set $\mathbb{Z} - (p)$. The prime ideals of $\mathbb{Z}_{(p)}$ correspond to the ideals of $\mathbb{Z}$ contained in (p) and there two of them (p) and (0). Thus,

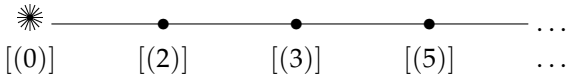
  $$\mathrm{Spec}(\mathbb{Z}_{(p)}) = \{[0], [(p)]\}.$$

  The point $[(p)]$ is closed. The point $[0]$ is an open set whose closure is the whole of $\mathrm{Spec}(\mathbb{Z}_{(p)})$.

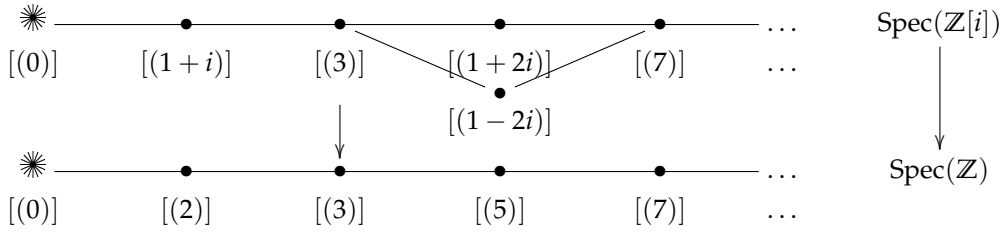
- Consider now $\mathrm{Spec}(\mathbb{Z})$. We know its points are $\{[0], [(2)], [(3)], [(5)], \dots\}$. Every point $[(p)]$ for p a prime number, is a closed point. The point $[(0)]$ is not a closed point; its closure is the whole of $\mathrm{Spec}(\mathbb{Z})$. Non-trivial closed sets are all of the form $V((n))$ where $n > 1$ is an integer. The points

in $V((n))$ correspond to the prime divisors of n. Every non-empty open set contains all but finitely many closed points (in particular, it contains $[0]$), and conversely. We draw $\mathrm{Spec}(\mathbb{Z})$ like that:



The situation for $\mathrm{Spec}(\mathbb{C}[x])$ (or, more generally, any PID) is quite similar. The point $\{0\}$ is dense and any proper closed subset is of the form $\{[(x-\alpha_1)],\dots,[(x-\alpha_n)]\}$, where the α_i are complex numbers. (For a general PID use irreducible elements, determined up to units.)

- The situations for rings of the form $\mathbb{Z}[i],\mathbb{Z}[\sqrt{2}],\mathbb{Z}[e^{2\pi i/7}]$ and so on is similar, but the precise enumeration of the prime ideals becomes a number-theoretic question. For example, for $\mathbb{Z}[i]$ we have closed points of the form (p) for every $p \equiv 3 \pmod 4$ and points of the form $[(x+yi)]$ and $[(x-yi)]$ for every prime $p \equiv 1 \pmod 4$, where x,y satisfy $x^2+y^2=p$; this is a theorem due to P. Fermat. An additional closed point is $[(1+i)]$. These are, in fact, all the closed points, but that requires proof. The only additional point of $\mathrm{Spec}(\mathbb{Z}[i])$ is $[(0)]$ and it's dense.



Proposition 5.2.5. *Let $f\colon R \to S$ be a homomorphism of rings. Then the induced map*

$$f^*\colon \mathrm{Spec}(S) \to \mathrm{Spec}(R)$$

is continuous.

Proof. We only need to calculate $(f^*)^{-1}(V(\mathfrak{a}))$ for an ideal $\mathfrak{a}$ of R. We claim that

$$(f^*)^{-1}(V(\mathfrak{a})) = V(\langle f(\mathfrak{a})\rangle).$$

(Here $\langle f(\mathfrak{a})\rangle$ is the ideal of S generated by the set $f(\mathfrak{a})$.) Indeed, if $\mathfrak{p}$ is an ideal of S that contains $f(\mathfrak{a})$ then $f^{-1}(\mathfrak{p}) \supseteq \mathfrak{a}$ and so $f^*[\mathfrak{p}] \in V(\mathfrak{a})$. That is, $\mathfrak{p} \in (f^*)^{-1}(V(\mathfrak{a}))$.

Conversely, if $\mathfrak{p} \in (f^*)^{-1}(V(\mathfrak{a}))$ then $f^*[\mathfrak{p}] = [f^{-1}(\mathfrak{p})] \in V(\mathfrak{a})$ so $f^{-1}(\mathfrak{p}) \supseteq \mathfrak{a}$ and it follows that $\mathfrak{p} \supseteq f(\mathfrak{a})$ and so $\mathfrak{p} \in V(\langle f(\mathfrak{a})\rangle)$. □

Corollary 5.2.6. *$R \mapsto \mathrm{Spec}(R)$ is a contravariant functor* **CommRings** $\to$ **Top**.

We have defined a topology on $\mathrm{Spec}(R)$ by specifying the closed sets. There is a particular kind of open sets that is very convenient to work with. Let $f \in R$ and define

$$D(f) = \{[\mathfrak{p}] : f \notin \mathfrak{p}\} = \mathrm{Spec}(R) - V((f)).$$

If we think about elements of R as functions on $\mathrm{Spec}(R)$ (and we shall soon see that this is precisely the case) then we can think about $f \pmod{\mathfrak{p}}$ as the *value* of f at the point $[\mathfrak{p}]$. Denote this value[12] $\bar{f}([\mathfrak{p}])$ then

$$\bar{f}([\mathfrak{p}]) = 0 \Leftrightarrow f \in \mathfrak{p}.$$

[12]We use $\bar{f}([\mathfrak{p}])$ and not just $f([\mathfrak{p}])$ because soon we will use the notation $f([\mathfrak{p}])$ to denote something differnt.

Under this interpretation of elements of R as functions, $D(f)$ is precisely the set where $f \neq 0$. The complement of $D(f)$ is $V((f))$.

For example, if $R = \mathbb{C}[x_1, \dots, x_n]$ and $f(x_1, \dots, x_n)$ is in R, then at the closed point $[\mathfrak{p}] = [(x_1 - \alpha_1, \dots, x_n - \alpha_n)]$ defined by an n-tuple of complex numbers $\alpha_1, \dots, \alpha_n$, the value of the polynomial f is classically $f(\alpha_1, \dots, \alpha_n)$, which is precisely $f \pmod{\mathfrak{p}}$.

The sets $D(f)$ are a basis for the topology, i.e. they are open and any open set is a union of the sets $D(f)$. Indeed, if $[\mathfrak{p}] \notin V(\mathfrak{a})$ then there is an element $f \in \mathfrak{a}$ such that $f \notin \mathfrak{p}$. Then we see that $D(f)$ is an open set that is disjoint from $V(\mathfrak{a})$ and contains $[\mathfrak{p}]$.

Proposition 5.2.7. *Let $\ell\colon R \to R[f^{-1}]$ be the homomorphism into the localization of R in the multiplicative set $\{f^n\}_{n\geq 0}$, $\ell(r) = \frac{r}{1}$. The function*

$$\ell^* : \mathrm{Spec}(R[f^{-1}]) \to \mathrm{Spec}(R)$$

is a homemorphism onto $D(f)$.

Proof. We have seen in Example 5.1.4 that this map is a bijection from $\mathrm{Spec}(R[f^{-1}])$ to $D(f)$, and Proposition 5.2.5 gives that this function ℓ^* is continuous. It remains to show that it is a closed map when $D(f)$ is provides with the induced topology.

Let $\mathfrak{a}$ be an ideal of $R[f^{-1}]$. Let I be a prime ideal of $R[f^{-1}]$. We claim that $I \supseteq \mathfrak{a} \Leftrightarrow I^c \supseteq \mathfrak{a}^c$. One direction is obvious. Suppose that $I^c \supseteq \mathfrak{a}^c$ then $I = I^{ce} \supseteq \mathfrak{a}^{ce}$. So it only remains to check that $\mathfrak{a}^{ce} \supseteq \mathfrak{a}$, and, indeed if $\frac{a}{s} \in \mathfrak{a}$ then $\frac{a}{1} \in \mathfrak{a}$ and also $\frac{a}{1} \in \mathfrak{a}^{ce}$. But then also $\frac{a}{s} \in \mathfrak{a}^{ce}$.

Let $\mathfrak{a}$ be an ideal of $R[f^{-1}]$ then $\ell^*(V(\mathfrak{a}))$ is the set

$$\{I^c : I \supseteq \mathfrak{a}, I \text{ prime}\} = \{J \lhd R \text{ prime}, f \notin J, J \supseteq \mathfrak{a}^c\} = D(f) \cap V(\mathfrak{a}^c),$$

which is a closed set of $D(f)$. □

Remark 5.2.8. To study the prime ideals of R containing in $\mathfrak{p}$ we may pass to the local ring $R_{\mathfrak{p}}$. Note that the ring homorphism $R \to R_{\mathfrak{p}}$ provides an injective continuous map

$$\mathrm{Spec}(R_{\mathfrak{p}}) \to \mathrm{Spec}(R),$$

whose image, in general is not an open set of $\mathrm{Spec}(R)$ (due to $R - \mathfrak{p}$ not being finitely generated as a semi-group in general). For a concrete example, consider $\mathfrak{p} = 0$ in $\mathbb{Z}$. The localization is $\mathbb{Q}$ and $\mathrm{Spec}(\mathbb{Q})$ has only one point $[0]$. Its image is the point $[0]$ of $\mathrm{Spec}(\mathbb{Z})$ that is not an open set, as its closure is not of the form $V(\mathfrak{a})$ for any ideal $\mathfrak{a} \lhd \mathbb{Z}$.

To study ideals containing $\mathfrak{p}$, we may pass to the integral domain $R/\mathfrak{p}$ and the ring homomorphism $R \to R/\mathfrak{p}$ induces an injective continuous map

$$\mathrm{Spec}(R/\mathfrak{p}) \to \mathrm{Spec}(R),$$

Whose image is the closed set $V(\mathfrak{p})$. It is a homeomorphism onto its image.

The methods complement each other, but the images of $\mathrm{Spec}(R_{\mathfrak{p}})$ and $\mathrm{Spec}(R/\mathfrak{p})$ do not complement each other.

Remark 5.2.9. Hochster (Prime ideal structure in commutative rings. TAMS 142 (1969), 43–60) gave a topological characterization of all topological spaces isomorphic to the spectrum of some commutative ring. Such spaces are the topological spaces X that have the following properties:

(1) X is a sober space, meaning any closed irreducible set is the closure of a unique point;
(2) X is quasi-compact, meaning any open cover has a finite subcover;
(3) the intersection of two quasi-compact open subsets is a quasi-compact open set;

(4) the quasi-compact open sets of X form a basis for its topology.

Let R be a ring and $n \geq 0$ an integer. One denotes

$$\mathbb{A}^n_R = \mathrm{Spec}(R[x_1, \ldots, x_n]),$$

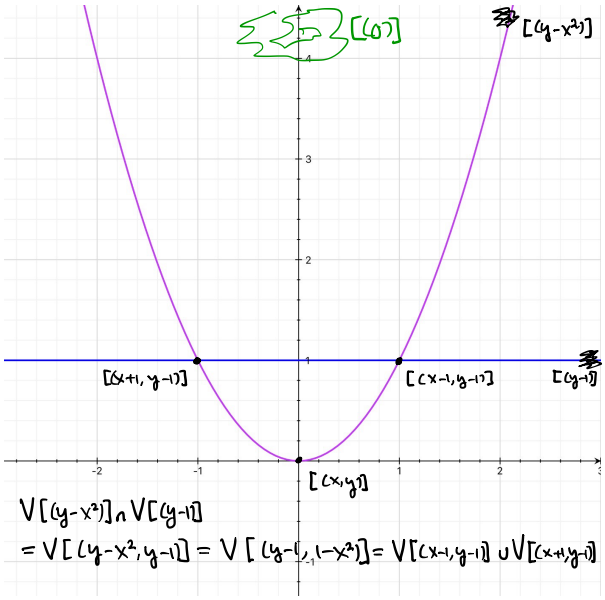
and calls it the n-dimensional **affine space over** R. Let us look more closely at two examples:

(1) **The complex affine line $\mathbb{A}^1_{\mathbb{C}} = \mathrm{Spec}(\mathbb{C}[x])$.** We have seen that the points of this space are $[0]$ and $[(x-\alpha)]$ for $\alpha \in \mathbb{C}$. The point $[0]$ is dense – its closure is the whole space. Every other point $[\mathfrak{p}]$ is a closed point, namely, the closure of the set $\{[\mathfrak{p}]\}$ is itself.

Every non-zero ideal $\mathfrak{a}$ is of the form $(f(x))$ and the set $V(\mathfrak{a})$ consists of the points $[(x-\alpha)]$ such that $(x-\alpha)|f(x)$, namely, of the points $[(x-\alpha)]$ such that $f(\alpha) = 0$. So, closed sets correspond to zeros of polynomials. One can also check that the complement of $V((f))$ is $D(f)$ consisting of the points where f doesn't vanish.

(2) **The complex affine plane $\mathbb{A}^2_{\mathbb{C}} = \mathrm{Spec}(\mathbb{C}[x,y])$.** The description requires two difficult results that are a special case of Hilbert's Nullstellensatz and Krull's Hauptidealsatz. They say in this case that
   (a) every maximal ideal of $\mathbb{C}[x,y]$ is of the form $(x-\alpha, y-\beta)$, and
   (b) besides (0) every other prime ideal of $\mathbb{C}[x,y]$ is of the form $(f(x,y))$ where $f(x,y) \in \mathbb{C}[x,y]$ is an irreducible polynomial, determined uniquely up to multiplication by a scalar.

The closure of the point $[(0)]$ is the whole $\mathbb{A}^2_{\mathbb{C}}$. If $f(x,y)$ is irreducible, the closure of a point $[(f(x,y))]$ is the point itself together with all the points $[(x-\alpha, y-\beta)]$ such that $f(\alpha,\beta) = 0$. The points $[(x-\alpha, y-\beta)]$ are closed. It requires additional commutative algebra to conclude that every closed set of $\mathbb{A}^2_{\mathbb{C}}$ is a finite union of these basic closed sets. For a polynomial $f(x,y) \neq 0$, the set $D(f)$ is the open set containing all points $[(x-\alpha, y-\beta)]$ such that $f(\alpha,\beta) \neq 0$ and all points $[(g(x,y))]$ such that g is irreducible and $g \nmid f$. Namely, the complement of $D(f)$ is the union of the finitely many closed sets $V([(g(x,y))])$ where $g(x,y)$ varies over all irreducible factors of $f(x,y)$.



5.3. $\mathrm{Spec}(R)$ **as a locally ringed space.** The additional structure one puts on $\mathrm{Spec}(R)$ is that of a sheaf. We therefore begin by introducing this notion.

5.3.1. *Sheaves.* Let X be a topological space. A **sheaf** $\mathcal{O}$ of abelian groups (resp., commutative rings) on X is the following data:

(1) (*values*) For each open set U an abelian group (resp. commutative ring) $\mathcal{O}(U)$, with $\mathcal{O}(\emptyset) = 0$.
(2) (*restriction maps*) For each inclusion $V \subseteq U$ a homomorphism $res_{UV}\colon \mathcal{O}(U) \to \mathcal{O}(V)$ of groups (resp., rings) such that (i) $res_{UU} = Id$ and (ii) for $V \subseteq U \subseteq W$ we have $res_{UV} \circ res_{WU} = res_{WV}$.
(3) (*locally zero is zero*) If $U = \cup U_i$ and $s \in \mathcal{O}(U)$ is such that $res_{UU_i}(s) = 0$ for all i, then $s = 0$.
(4) (*local data can be glued*) If $U = \cup U_i$ and $s_i \in \mathcal{O}(U_i)$ are elements such that for all i, j, $res_{U_i, U_i \cap U_j}(s_i) = res_{U_j, U_i \cap U_j}(s_j)$ then there exists $s \in \mathcal{O}(U)$ such that $res_{UU_i}(s) = s_i$ for all i.

A **ringed space** X is a topological space with a sheaf of rings $\mathcal{O}$ (or $\mathcal{O}_X$, if we want to specify the space). Given a point $x \in X$ we can form the ring of **germs of functions** (also called the **stalk** of the sheaf) at x by

$$\mathcal{O}_{X,x} = \varinjlim_{x \in U} \mathcal{O}_X(U),$$

the limit taken over all open sets U containing x and the homomorphisms are the restriction maps. As the index set here is directed, we can think about an element of this ring as a pair

$$(U, f)$$

consisting of an open set U containing x and $f \in \mathcal{O}_X(U)$; two pairs $(U, f), (V, g)$ are considered the same element in $\mathcal{O}_{X,x}$ if $res_{U,W}(f) = res_{V,W}(g)$ for some open set $W \subset U \cap V$ that contains x. We will also write the last equation more transparently as $f|_W = g|_W$. In this language, for example,

$$(U_1, f_1) + (U_2, f_2) = (U_1 \cap U_2, f_1|_{U_1 \cap U_2} + f_2|_{U_1 \cap U_2}).$$

In spite of the terminology, the ring $\mathcal{O}(U)$ does not a priori consist of functions on U and so certain constructions concerning sheaves have to be carried out without using an interpretation of the elements of $\mathcal{O}(U)$ as functions.

In general, the ring $\mathcal{O}_{X,x}$ need not be a local ring; recall that a local ring R, by definition, is a ring that has a unique maximal ideal. This need not be the case for $\mathcal{O}_{X,x}$. A ringed space $(X, \mathcal{O}_X)$ is called a **locally ringed space** if the sheaf of rings $\mathcal{O}_X$ has the property that the rings $\mathcal{O}_{X,x}$ are local rings for all $x \in X$. We then refer to $\mathcal{O}_{X,x}$ also as the **local ring of** x.

Example 5.3.1. Let $U \subset \mathbb{R}^n$ be an open set. Then, defining for an open set $V \subseteq U$,

$$\mathcal{O}(V) = \{f\colon V \to \mathbb{R}, \text{ continuous}\},$$

makes U into a locally ringed space. The maximal ideal of the ring $\mathcal{O}_x$ are all the pairs (V, f) such that $f(x) = 0$. If we choose to define

$$\mathcal{O}(V) = \{f\colon V \to \mathbb{R}, f \text{ is } C^\infty\},$$

we would get a different locally ringed space (with the same underlying topological space).

Example 5.3.2. A Riemann surface with its sheaf of analytic functions is a locally ringed space.

Example 5.3.3. Let a be a symbol and consider $\{a\}$ as a topological space with one point. Define a sheaf on it by $\mathcal{O}(\{a\}) = R$, where R is some commutative ring. Then $\mathcal{O}_a = R$. If R is not a local ring this is a ringed space that is not a locally ringed space.

5.3.2. *The sheaf on* $\mathrm{Spec}(R)$. We wish to make $X = \mathrm{Spec}(R)$ into a locally ringed space. We first note that following: for $[\mathfrak{p}] \in \mathrm{Spec}(R)$ the ring $R_{\mathfrak{p}}$ is a local ring; its unique maximal ideal is the localization of $\mathfrak{p}$, equal to $\mathfrak{p}R_{\mathfrak{p}}$. Thus, we wish to define a sheaf of rings $\mathcal{O}$ on $\mathrm{Spec}(R)$ with the property that the local ring $\mathcal{O}_{[\mathfrak{p}]} = R_{\mathfrak{p}}$. We do that as follows:

For an open set U let $\mathcal{O}(U)$ be functions f on U with the property that for all $[\mathfrak{p}] \in U$ we have $f(\mathfrak{p}) \in R_{\mathfrak{p}}$ and such that f is locally a ratio of two elements of R. Namely, for each $[\mathfrak{p}] \in U$ exists an open set $V \subseteq U$ containing $[\mathfrak{p}]$ and elements r, s of R such that $s \notin \mathfrak{q}$ for all points $[\mathfrak{q}] \in V$ and such that $f = \frac{r}{s}$ in $R_{\mathfrak{q}}$ for all $[\mathfrak{q}] \in V$.

Given two functions f, g, in $\mathcal{O}(U)$ we define $f + g$ by $(f + g)([\mathfrak{p}]) = f([\mathfrak{p}]) + g([\mathfrak{p}])$, and similarly for products. One needs to check that the sum and product, thus defined, are also "locally fractions" and that is easy. This gives $\mathcal{O}(U)$ a ring structure. The natural restriction maps for $V \subseteq U$ are indeed ring homomorphisms $\mathcal{O}(U) \to \mathcal{O}(V)$. Given an open cover $U = \cup_i U_i$, and $f \in \mathcal{O}(U)$ such that $f|_{U_i} = 0$, clearly $f = 0$. Given functions $f_i \in \mathcal{O}(U_i)$ such that $f_i|_{U_i \cap U_j} = f_j|_{U_i \cap U_j}$ we define a function f on U by $f([\mathfrak{q}]) = f_i([\mathfrak{q}])$ if $[\mathfrak{q}] \in U_i$. This is well-defined. The only further verification required is that f is locally a fraction. It is enough to show that $f|_{U_i}$ is locally a fraction, but that is clear because $f|_{U_i} = f_i$.

We therefore get a sheaf of rings on $\mathrm{Spec}(R)$ making it into a ringed space. To show it is a locally ringed space we prove the following lemma.

Lemma 5.3.4. *There is a natural isomorphism* $\mathcal{O}_{[\mathfrak{p}]} \cong R_{\mathfrak{p}}$.

Proof. Any element of $\mathcal{O}_{[\mathfrak{p}]}$ has a representative (U, f), where U is an open set containing $\mathfrak{p}$, and to construct the isomorphism we associate it the value $f([\mathfrak{p}]) \in R_{\mathfrak{p}}$. This is independent of the representative and provides a ring homomorphism $\mathcal{O}_{[\mathfrak{p}]} \to R_{\mathfrak{p}}$.

Suppose that (U, f) is mapped to zero. There is an open set V containing $[\mathfrak{p}]$ such that on V the function f is a fraction $\frac{r}{s}$. By definition, there is some $s_1 \in R - \mathfrak{p}$ such that $s_1 r = 0$. The open set $D(s_1) \cap V$ contains $[\mathfrak{p}]$ and on it f is zero: for each point $[\mathfrak{q}]$ in it, $f = \frac{r}{s}$ is also zero in $R_{\mathfrak{q}}$ because $s_1 r = 0$ and $s_1 \notin \mathfrak{q}$. So we have $(U, f) = (D(s_1) \cap V, f|_{D(s_1) \cap V}) = (D(s_1) \cap V, 0) = 0$.

Finally, our map is surjective. Let $\frac{r}{s} \in R_{\mathfrak{p}}$, then $(D(s), \frac{r}{s})$ is a well-defined element of $\mathcal{O}_{[\mathfrak{p}]}$ mapping to $\frac{r}{s}$ under our ring homomorphism. □

Remark 5.3.5. One can show that

$$\mathcal{O}(D(f)) = R[f^{-1}],$$

the localization of R in the multiplicative set $\{1, f, f^2, \dots\}$ (See *Hartshorne, Algebraic Geometry*, Graduate Texts in Mathematics 52, Springer). This requires some effort and we will not prove this here. However, we will use it, and the next proposition that strengthens it a bit, in the sequel. Note that, in particular, taking $f = 1$ we find

$$\mathcal{O}(\mathrm{Spec}(R)) = R.$$

Remark 5.3.6. Let $(X, \mathcal{O}_X)$ be a locally ringed space and $U \subseteq X$ an open subset. The **induced sheaf** $\mathcal{O}_X|_U$ is defined as follows: for $V \subseteq U$ open let

$$V \mapsto \mathcal{O}_X(V).$$

It is immediate that this is a sheaf of rings making $(U, \mathcal{O}_X|_U)$ into a locally ringed space (with the same local rings!). The next proposition is left as an exercise.

Proposition 5.3.7. *Let $f \in \mathrm{Spec}(R)$, then $D(f)$ with the induced sheaf $\mathcal{O}_X|_{D(f)}$ is isomorphic to $\mathrm{Spec}(R[f^{-1}])$.*

5.3.3. *Morphisms of locally ringed spaces.* It is natural to expect at this point that a ring homomorphism $f\colon R \to S$ would produce a morphism of locally ringed spaces

$$f^*\colon \mathrm{Spec}(S) \to \mathrm{Spec}(R).$$

This is correct, but we haven't yet defined what that means!

To begin with we define a morphism of sheaves: Let X be a topological space and $\mathscr{F},\mathscr{G}$ two sheaves of commutative rings on X. A **morphism of sheaves**,

$$h\colon \mathscr{F} \to \mathscr{G},$$

is a collection of ring homomorphisms

$$h_U\colon \mathscr{F}(U) \to \mathscr{G}(U),$$

such that for all $V \subseteq U$ the following diagram commutes:

$$\begin{array}{ccc} \mathscr{F}(U) & \xrightarrow{h_U} & \mathscr{G}(U) \\ \downarrow{\scriptstyle res_{UV}} & & \downarrow{\scriptstyle res_{UV}} \\ \mathscr{F}(V) & \xrightarrow{h_V} & \mathscr{G}(V) \end{array}$$

If we think of $\mathscr{F}$ and $\mathscr{G}$ as functors from the poset of open sets of X to commutative rings, h is just a natural transformation of functors.

Let $(X,\mathcal{O}_X),(Y,\mathcal{O}_Y)$ be ringed spaces. Suppose that $f\colon X \to Y$ is a continuous map. Then, we get a new sheaf of rings on Y, denoted $f_*\mathcal{O}_X$ defined by

$$f_*\mathcal{O}_X(U) = \mathcal{O}_X(f^{-1}(U)).$$

We leave the verification that this is a sheaf of rings as an exercise. A priori there is no relation between $\mathcal{O}_Y(U)$ and $\mathcal{O}_X(f^{-1}(U))$ as the sheaves are not guaranteed to be sheaves of functions on the space X or Y (although in almost any application they are...). Thus, we also specify a homomorphism of sheaves

$$f^\sharp : \mathcal{O}_Y \to f_*\mathcal{O}_X.$$

Heuristically, $f^\sharp$ tells us how to pull back "functions" from Y to X. Indeed, in many situations the map $f^\sharp$ is so obvious that it doesn't require mention at all. A **morphism of ringed spaces** is thus such a pair:

$$(f,f^\sharp)\colon (X,\mathcal{O}_X) \to (Y,\mathcal{O}_Y),$$

where $f\colon X \to Y$ is a continuous map and $f^\sharp\colon \mathcal{O}_Y \to f_*\mathcal{O}_X$ is a morphism of sheaves.

For example, suppose that X,Y are topological spaces and $\mathcal{O}_X,\mathcal{O}_Y$ are the sheaves of real-valued continuous functions on X,Y, respectively. Then we can let $f^\sharp$ be composition with f. For every open set $U \subset Y$ and $g\colon U \to \mathbb{R}$ a continuous map, let $f^\sharp(g) := g \circ f$; it is a continuous function on $f^{-1}(U)$. That is, $f^\sharp(g) \in f_*\mathcal{O}_X(U)$.

We need to make further conditions on such morphisms. Otherwise we cannot prove the main theorem we are after. Returning to the general case, let $x \in X$ and $y = f(x)$. We claim that we have a natural map

$$\mathcal{O}_{Y,y} \to \mathcal{O}_{X,x}.$$

Let I be the poset of open sets of Y that contain the point y, and let J the poset of open sets of X that contain the point x. Then

$$\mathcal{O}_{Y,y} = \varinjlim_{U\in I} \mathcal{O}_Y(U), \qquad \mathcal{O}_{X,x} = \varinjlim_{U\in J} \mathcal{O}_X(U).$$

The homomorphisms $f^\sharp$ then provide maps

$$\mathcal{O}_{Y,y} = \varinjlim_{U\in I} \mathcal{O}_Y(U) \xrightarrow{f^\sharp} \varinjlim_{U\in I} \mathcal{O}_X(f^{-1}(U)) \longrightarrow \varinjlim_{V\in J} \mathcal{O}_X(V) = \mathcal{O}_{X,x}.$$

(This is perhaps easiest to check if we think about the elements of the local rings as equivalence classes of pairs (U,g).)

Suppose that $(X,\mathcal{O}_X),(Y,\mathcal{O}_Y)$, are locally ringed spaces and

$$(f,f^\sharp) : (X,\mathcal{O}_X) \to (Y,\mathcal{O}_Y)$$

is a morphism of ringed spaces. To be a **morphism of locally ringed spaces** we make the additional requirement that for all $x \in X$ the induced ring homomorphism

$$f^\sharp : \mathcal{O}_{Y,f(x)} \to \mathcal{O}_{X,x}$$

is a **local homomorphism**. Namely, that

$$f^{\sharp,-1}(\mathfrak{m}_x) = \mathfrak{m}_{f(x)},$$

where $\mathfrak{m}_x$ (resp. $\mathfrak{m}_{f(x)}$) is the maximal ideal of $\mathcal{O}_{X,x}$ (resp. $\mathcal{O}_{Y,f(x)}$).

Note that because $\mathcal{O}_{Y,f(x)}$ is a local ring and $f^{\sharp,-1}(\mathfrak{m}_x)$ is a prime ideal, we always have $f^{\sharp,-1}(\mathfrak{m}_x) \subseteq \mathfrak{m}_{f(x)}$. To say that we have equality is equivalent to saying that $f^\sharp(\mathfrak{m}_{f(x)}) \subset \mathfrak{m}_x$. We can phrase that as saying that every function on Y that vanished at $f(x)$ is pulled-back to a function on X that vanishes at x; this is certainly a natural condition from a geometric point of view!

5.4. **An equivalence of categories.** By definition, an **affine scheme** is a locally ringed space isomorphic as a locally ringed space to $(\mathrm{Spec}(R),\mathcal{O})$ for some ring R. Morphisms of affine schemes are morphisms of locally ringed spaces. Thus, affine schemes are a full subcategory **AffSch** of the category of locally ringed spaces. Our main theorem in this part of the course is the following.

Theorem 5.4.1. *The category of affine schemes is anti-equivalent to the category* **CommRings** *of commutative rings.*

Proof. We define a functor

$$\mathbf{CommRings} \to \mathbf{AffSch}$$

by sending a ring R to $\mathrm{Spec}(R)$. Let $f\colon R \to S$ be a homomorphism of rings. We associate to it a morphism

$$(f^*,f^\sharp)\colon \mathrm{Spec}(S) \to \mathrm{Spec}(R).$$

We have already defined f^* as $f^*([\mathfrak{p}]) = [f^{-1}(\mathfrak{p})]$, and we showed it is a continuous map of topological spaces. To ease notation, write

$$X = \mathrm{Spec}(S), \;\; Y = \mathrm{Spec}(R).$$

We need to define the map

$$f^\sharp : \mathcal{O}_Y \to f_*\mathcal{O}_X.$$

Let U be an open set of $Y = \mathrm{Spec}(R)$ and $g \in \mathcal{O}_Y(U)$. Let $\mathfrak{q} \in (f^*)^{-1}(U)$, which means that $\mathfrak{p} = f^{-1}(\mathfrak{q}) \in U$. Then, by definition, $g([\mathfrak{p}]) \in R_\mathfrak{p}$. But, there is a canonical ring homomorphism $R_\mathfrak{p} \to S_\mathfrak{q}$ induced from

the homomorphism $R \to S$. It is simply given by $r/s \mapsto f(r)/f(s)$. Via this homomorphism we may view $g([\mathfrak{p}])$ as an element of $S_{\mathfrak{q}}$ that we shall call $f^\sharp(g)([\mathfrak{q}])$.

By definition $f^\sharp(g)$ is a function on $f^{-1}(U)$ such that $f^\sharp(g)([\mathfrak{q}]) \in S_{\mathfrak{q}}$ for all points $[\mathfrak{q}] \in U$. We need to show that this function is locally a fraction. Let $\mathfrak{q}, \mathfrak{p}$ be as before then there is an open set $V \subseteq U$, containing $\mathfrak{p}$ such that on V we have $g = r/s$, where s does not belong to any prime ideal $\mathfrak{p}' \in V$. Note then that if $[\mathfrak{q}']$ is such that $f^*([\mathfrak{q}']) = [\mathfrak{p}']$ then $f(s) \notin \mathfrak{q}'$. Under our interpretation of $f^\sharp(g)([\mathfrak{q}'])$, we have that $f^\sharp(g)([\mathfrak{q}'])$ is the image of $g([\mathfrak{p}])$ under $R_{\mathfrak{p}'} \to S_{\mathfrak{q}'}$, namely, the image of r/s, which is $f(r)/f(s)$. Thus, $f^\sharp(g)$ is represented on $f^{-1}(V)$ by $f(r)/f(s)$.

We note that $f^\sharp$ is a morphism of locally ringed spaces. In fact, the induced map on local rings is precisely

$$f\colon R_{\mathfrak{p}} \to S_{\mathfrak{q}},$$

and not only $f^{-1}(\mathfrak{q}) = \mathfrak{p}$ but also $f^{-1}(\mathfrak{q}^e) = \mathfrak{p}^e$, where "e" denotes extended ideals.

It is now easy to check that we got a contravariant functor $\mathbf{CommRings} \to \mathbf{AffSch}$. By definition it is essentially surjective. It remains to show it is fully-faithful. Taking the open set $U = \mathrm{Spec}(R)$ and $g \in R$ viewed as an element of $\mathcal{O}_X(U)$, per definition the associated element in $\mathrm{Spec}(S) = (f^*)^{-1}(U)$ is simply $f(g)$. Thus, the functor is faithful. It remains to show it is full.

For that we use the following argument, left as an exercise: *if two morphisms $X \to Y$ of locally ringed spaces agree on a collection of open sets that form a basis for Y and their pre-images in X then they are the same morphism.*

We use this principle for the collection $\{D(h) : h \in R\}$ that covers Y. Let $D(h)$ be a basic open set of $Y = \mathrm{Spec}(R)$. Then $(f^*)^{-1}(D(h)) = D(f(h))$ is a basic open set of $\mathrm{Spec}(S)$. We have

$$\mathcal{O}_X(D(h)) = R[h^{-1}].$$

In particular, every function g on $D(h)$ has a *global* representation as r/h^n on $D(h)$. From our definition, $f^\sharp(g) = f(r)/f(h)^n$.

Now, let $(F, F^\sharp)\colon X \to Y$ be any morphism of locally ringed spaces. Then, taking $U = Y = \mathrm{Spec}(R)$ we have a ring homomorphism

$$F^\sharp\colon R = \mathcal{O}_Y(Y) \to S = \mathcal{O}_X(X).$$

Denote this ring homomorphism f. We claim that

$$(f^*, f^\sharp) = (F, F^\sharp).$$

If so, we proved that the functor is full.

Let $\mathfrak{p}$ be a point of X. We have a commutative diagram:

$$\begin{array}{ccc} \mathcal{O}_Y(Y) = R & \xrightarrow{f=F^\sharp} & F_*\mathcal{O}_X(Y) = S \\ \downarrow & & \downarrow \\ \mathcal{O}_{Y,F(\mathfrak{p})} & \xrightarrow{F^\sharp} & \mathcal{O}_{X,\mathfrak{p}} \end{array}$$

Now, there is a unique way to extend $F^\sharp$ to the localization[13] and so, on the localized rings, we must have $F^\sharp = f$, too. Since the homomorphism of local rings is a local homomorphism, we conclude that in fact

$$F(\mathfrak{p}) = (f)^{-1}(\mathfrak{p}) = f^*(\mathfrak{p}).$$

[13] If s is an element of multiplicative set $1 = F^\sharp(1) = F^\sharp(\frac{s}{s}) = F^\sharp(\frac{s}{1}) \cdot F^\sharp(\frac{1}{s}) = \frac{F^\sharp(s)}{1} \cdot F^\sharp(\frac{1}{s})$. This determines $F^\sharp(\frac{1}{s})$ and more generally $F^\sharp(\frac{r}{s})$.

It follows that for the topological maps we have $F = f^*$. It remains to check that on every open set U of X the map $f^\sharp$ agrees with $F^\sharp$. But it is enough to check that on the basis open sets $U = D(f)$. In this case the rings are again localizations and the argument goes as for the local rings. □

Corollary 5.4.2. *Pull-back (fibre product) exists in the category of affine schemes.*

Proof. Given a diagram in **AffSch**,

$$\begin{array}{ccc} & & \mathrm{Spec}(R) \\ & & \downarrow \\ \mathrm{Spec}(S) & \longrightarrow & \mathrm{Spec}(A), \end{array}$$

we have the dual diagram in **CommRings**:

$$\begin{array}{ccc} & & R \\ & & \uparrow \\ S & \longleftarrow & A \end{array}$$

We have a *push-out* in **CommRings**:[14]

$$\begin{array}{ccc} R \otimes_A S & \longleftarrow & R \\ \uparrow & & \uparrow \\ S & \longleftarrow & A \end{array}$$

It provides us with a *pull-back* in **AffSch**

$$\begin{array}{ccc} \mathrm{Spec}(R \otimes_A S) & \longrightarrow & \mathrm{Spec}(R) \\ \downarrow & & \downarrow \\ \mathrm{Spec}(S) & \longrightarrow & \mathrm{Spec}(A). \end{array}$$

□

Note an important point: the point set of $\mathrm{Spec}(R \otimes_A S)$ is not the fibre product of the point sets of $\mathrm{Spec}(R)$ and $\mathrm{Spec}(S)$ over $\mathrm{Spec}(A)$. For example take

$$A = \mathbb{C}, \quad R = \mathbb{C}[x], \quad S = \mathbb{C}[y].$$

Then, as sets

$$\mathrm{Spec}(\mathbb{C}) = [0], \quad \mathrm{Spec}(\mathbb{C}[x]) = \{[0]\} \cup \{[(x-\alpha)] : \alpha \in \mathbb{C}\}.$$

The fibre product as *sets* is therefore $\mathrm{Spec}(\mathbb{C}[x]) \times \mathrm{Spec}(\mathbb{C}[y])$. However, the point $[(x-y)] \in \mathrm{Spec}(\mathbb{C}[x] \otimes_{\mathbb{C}} \mathbb{C}[y]) = \mathrm{Spec}(\mathbb{C}[x,y])$ is not in this product.

As a matter of notation, given two affine schemes X, Y with morphisms $X \to Z \leftarrow Y$ we will denote the pullback of the diagram, the **fibre product of** X **and** Y **over** Z by

$$X \times_Z Y.$$

[14] It's a hard type-setting decision whether to use this diagram or the diagram

$$\begin{array}{ccc} A & \longrightarrow & R \\ \downarrow & & \downarrow \\ S & \longrightarrow & R \otimes_A S. \end{array}$$

6. Integral elements and integral extensions

6.1. **Integral elements.** Let $A \subseteq B$ be commutative rings. An element $b \in B$ is call **integral** over A if b solves a monic polynomial in $A[x]$.

Example 6.1.1. (1) If A and B are fields then "integral over A" is the same as "algebraic over A".

(2) Suppose that A is a UFD and $B = \mathrm{Frac}(A)$. Then b is integral over A if and only if $b \in A$. (So, for example, a rational number is integral over $\mathbb{Z}$ if and only if it is an integer.)
Proof. Every $b \in A$ solves $x - b = 0$ and so elements of A are always integral over A (for any ring A). Suppose $\alpha/\beta \in B$, $\alpha, \beta \in A, \beta \neq 0$. We may assume $\gcd(\alpha, \beta) = 1$. Suppose that α/β solves

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in A[x].$$

Then,

$$\alpha^n = -a_{n-1}\alpha^{n-1}\beta - \cdots - a_0\beta^n.$$

As β divides the right hand side, β divides α^n. As $\gcd(\alpha, \beta) = 1$ this means $\beta \in A^\times$ and so $\alpha/\beta \in A$.

(3) Let $\alpha \in \mathbb{C}$ be algebraic over $\mathbb{Q}$ then for some positive integer m, $m\alpha$ is integral over $\mathbb{Z}$.
Indeed, α solve some monic polynomial

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Q}[x],$$

so $m\alpha$ solves

$$x^n + ma_{n-1}x^{n-1} + \cdots + m^n a_0,$$

which for suitable $m > 0$ has integral coefficients.

Theorem 6.1.2. *Let $A \subseteq B$ be commutative rings and let $b \in B$. The following are equivalent:*

(1) b is integral over A.
(2) $A[b]$ is a finitely-generated A-module.
(3) $A[b] \subseteq M \subseteq B$ for some subring M of B that is a finitely-generated A-module.
(4) There is a faithful $A[b]$-module K such that K is a finitely generated A-module.

Proof.
(1) $\Rightarrow$ (2). Suppose b solves $x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in A[x]$. Then

$$b^n = -a_{n-1}b^{n-1} - \cdots - a_0 \in A + Ab + \cdots + Ab^{n-1}.$$

Arguing inductively one finds

$$b^N \in A + Ab + \cdots + Ab^{n-1}, \quad \forall N \geq n.$$

Therefore,

$$A[b] = A + Ab + Ab^2 + \cdots + Ab^N + \cdots = A + Ab + \cdots + Ab^{n-1}$$

is finitely generated over A.

(2) $\Rightarrow$ (3). Take $M = A[b]$.

(3) $\Rightarrow$ (4). Take $K = M$. It is a faithful $A[b]$-module because for any $c \in A[b]$, $c \neq 0$, we have $c = c \cdot 1_A \in K$ a non-zero element.

$(4) \Rightarrow (1)$. This is the most substantial part of the theorem. Let K be a faithful $A[b]$-module, finitely generated over A. Say,

$$K = Ac_1 + \cdots + Ac_n,$$

for some $c_i \in K$. Let $\varphi_b \colon K \to K$ be the multiplication-by-b map

$$\varphi_b(k) = bk, \quad k \in K.$$

Then, there are $a_{ij} \in A$ (not necessarily unique) such that

$$\varphi_b(c_i) = \sum_{j=1}^{n} a_{ij} c_j.$$

Consider the matrix

$$T = \begin{pmatrix} b & 0 & & 0 \\ 0 & b & & \\ & & \ddots & \\ 0 & & & b \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & & \\ & & \ddots & \\ a_{n1} & & & a_{nn} \end{pmatrix}.$$

We make no claim that T is a well defined map on K, but it does satisfy

$$T \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Now, T has entries in the ring B, in fact in $A[b]$, and using the definition of the adjoint matrix $Adj(T)$ we have

$$Adj(T) \cdot T = \det(T) \cdot I_n.$$

Indeed, such an identity holds when the entries of T are free variables and we can then substitute for them any values in any ring. Therefore,

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = Adj(T)(T \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}) = (Adj(T) \cdot T) \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \det(T) \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

That means that $\det(T)c_i = 0$ for all i and so $\det(T)k = 0$ for all $k \in K$. But, as K is faithful over $A[b]$, that implies $\det(T) = 0$. On the other hand, $\det(T) = f(b)$, where $f(x)$ is the characteristic polynomial of the matrix $xI_n - (a_{ij})$, which is a monic polynomail in $A[x]$. Therefore b is integral over A. □

Let $A \subseteq B$ be rings. The collection of elements of B that are integral over A is called the **integral closure** of A in B. If it is equal to A, we say A is **integrally closed** in B. If it is equal to B, we say $A \subseteq B$ is an **integral extension**.

Theorem 6.1.3. *The integral closure of A in B is a subring of B containing A. It is integrally closed in B.*

Proof. Let $b_1, b_2 \in B$ be integral over A. Then $A[b_1]$ and $A[b_2]$ are finitely generated over A and therefore $A[b_1, b_2]$ is finitely generated over A too. Since $A[-b_1], A[b_1 + b_2], A[b_1 b_2]$ are all contained in $A[b_1, b_2]$, we may apply part (3) of Theorem 6.1.2 with $M = A[b_1, b_2]$ and deduce that $-b_1, b_1 + b_2, b_1 b_2$ are integral over A too. It follows that the integral closure R of A in B is a subring of B.

We show now that R is integrally closed in B. Let $b \in B$ be integral over R. Then b solves some

$$x^n + r_{n-1}x^{n-1} + \cdots + r_0 \in R[x].$$

Thus, b is integral over the subring $A[r_0,\dots,r_{n-1}]$ and it follows that $A[r_0,\dots,r_{n-1},b]$ is finitely generated over $A[r_0,\dots,r_{n-1}]$. But, $A[r_0,\dots,r_{n-1}]$ is finitely generated over A, because each r_i is integral over A. Hence, $A[r_0,\dots,r_{n-1},b]$ is finitely generated over A. We have the following situation

$$A[b] \subseteq M = A[r_0,\dots,r_{n-1},b] \subseteq B.$$

Applying Theorem 6.1.2 again we conclude that b is integral over A. □

The following lemma is left as an exercise.

Lemma 6.1.4. *Let $A \subseteq B \subseteq C$ be commutative rings and let R be the integral closure of A in B, S the integral closure of A in C. Then S is also the integral closure of R in C.*

Example 6.1.5. Let $A = \mathbb{C}[x,y]/(x^3 - y^2)$, corresponding to a cuspidal curve $\mathscr{C}$. Let $t = y/x$. Note that $t^2 = y^2/x^2 = x$ and so $A \subset \mathbb{C}(t)$. It follows that A is an integral domain and $\mathbb{C}(t) = \mathrm{Frac}(A)$. Note that

$$A \subset \mathbb{C}[t] \subset \mathbb{C}(t).$$

The integral closure of A in $\mathbb{C}(t)$ contains $\mathbb{C}[t]$ because t solves the monic polynomial equation $u^2 - x$ with coefficients in A (u is the variable). By the previous lemma, the integral closure of A in $\mathbb{C}(t)$ is the same as the integral closure of $\mathbb{C}[t]$ in $\mathbb{C}(t)$, which is just $\mathbb{C}[t]$, because $\mathbb{C}[t]$ is a UFD. That is, the integral closure of A in $\mathbb{C}(t)$ is $\mathbb{C}[t]$.

Geometrically the curve corresponding to $\mathbb{C}[t]$ is $\mathrm{Spec}(\mathbb{C}[t])$ which is the affine line. The inclusion homomorphism $A \to \mathbb{C}[t]$ provides us with a map

$$\mathbb{A} = \mathrm{Spec}(\mathbb{C}[t]) \to \mathscr{C} = \mathrm{Spec}(A).$$

The theory we will develop in the sequel will allow us a very fine understanding of such maps. From a geometric point of view we managed to resolve the singularities of $\mathscr{C}$ by passing to a curve that is "very close" to it – they have the same field of rational functions $\mathbb{C}(t)$ and we managed to do that by passing to an integral closure – a very canonical construction. It turns out that this is a general method, called **normalization**, that is able to resolve the singularities of any curve and in general simplify the singularities of any variety (though not to resolve them completely).

6.2. **The case of number fields.**

The results of the previous section show that the elements of $\overline{\mathbb{Q}}$ that are integral over $\mathbb{Z}$, called **algebraic integers** form a ring $\mathcal{O}_{\overline{\mathbb{Q}}}$, called the ring of algebraic integers. If $K \subseteq \overline{\mathbb{Q}}$ is a **number field**, that is if K is a finite extension of $\mathbb{Q}$, then

$$\mathcal{O}_K := \mathcal{O}_{\overline{\mathbb{Q}}} \cap K$$

is the integral closure of $\mathbb{Z}$ in K, which is a subring of K called the **ring of algebraic integers** in K. It is one of the main objects of study of number theory.

Lemma 6.2.1. *Let $\alpha \in \overline{\mathbb{Q}}$. Then $\alpha \in \mathcal{O}_{\overline{\mathbb{Q}}}$ if and only if the minimal polynomial of α has integer coefficients.*

Proof. One direction is clear (the minimal polynomial is monic by definition). Assume then that α is an algebraic integer and let $f(x) \in \mathbb{Q}[x]$ be the monic minimal polynomial of α over $\mathbb{Q}$, and let $g(x) \in \mathbb{Z}[x]$ be a monic polynomial such that $g(\alpha) = 0$. Then $f(x) | g(x)$ in $\mathbb{Q}[x]$. This implies $f(x) \in \mathbb{Z}[x]$ too. This follows

from Gauss' lemma,[15] but one can also argue that the roots of f in $\overline{\mathbb{Q}}$ are among the roots of g and hence are algebraic integers. As the coefficients of f are symmetric functions in its roots, the coefficients of f are algebraic integers that also lie in $\mathbb{Q}$. Since $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$, the coefficients of f are integers. $\square$

In general, given a number field K, even if K has a nice description, calculating $\mathcal{O}_K$ is a non-trivial matter. Even the simplest examples such as $K = \mathbb{Q}(\zeta_n)$, where $\zeta_n = e^{2\pi i/n}$, require some theory and ingenuity (incidentally, the answer is $\mathcal{O}_K = \mathbb{Z}[\zeta_n]$). The examples we give are misleading in that the answer is very simple; in general, there is no need for $\mathcal{O}_K$ to be of the form $\mathbb{Z}[\alpha]$ and several generators may be needed. However, the situation is manageable for quadratic extensions of $\mathbb{Q}$.

Example 6.2.2. Let $d \neq 0,1$ be a square free integer and $K = \mathbb{Q}(\sqrt{d})$. Then

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{d}] & d \equiv 2,3 \pmod 4; \\ \mathbb{Z}[\frac{1+\sqrt{d}}{2}] & d \equiv 1 \pmod 4. \end{cases}$$

Proof. First note that when $d \equiv 1 \pmod 4$, $\mathbb{Z}[\frac{1+\sqrt{d}}{2}] \supset \mathbb{Z}[\sqrt{d}]$ and since $\frac{1+\sqrt{d}}{2}$ satisfies the polynomial $x^2 - x + \frac{(1-d)}{4}$, we can express $(\frac{1+\sqrt{d}}{2})^2 = \frac{(d-1)}{4} + \frac{1+\sqrt{d}}{2}$, which gives

$$\mathbb{Z}[\frac{1+\sqrt{d}}{2}] = \mathbb{Z} + \mathbb{Z} \cdot \frac{1+\sqrt{d}}{2},$$

and that every element of $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ is integral over $\mathbb{Z}$.

Now, a general element of K has the form $a + b\sqrt{d}$, $a, b \in \mathbb{Q}$. If $b = 0$ this element is integral if and only if $a \in \mathbb{Z}$. Assume that $b \neq 0$. The minimal polynomial of $a + b\sqrt{d}$ is

$$(x - (a + b\sqrt{d}))(x - (a - b\sqrt{d})) = x^2 - 2ax + (a^2 - db^2).$$

So $a + b\sqrt{d}$ is integral over $\mathbb{Z}$ if and only if $2a$ and $a^2 - db^2$ are integers.

If $a \in \mathbb{Z}$ then $db^2 \in \mathbb{Z}$ and since d is square free, $b \in \mathbb{Z}$. If $a = \alpha/2$ where α is an odd integer, write $b = \beta/2$ to find that

$$\alpha^2 - d\beta^2 \in 4\mathbb{Z}.$$

This implies that $\beta \in \mathbb{Z}$. Reducing modulo 4 we get $\alpha^2 - d\beta^2 \equiv 1 - d\beta^2 \equiv 0 \pmod 4$, which implies that $d \equiv 1 \pmod 4$. Thus, if $d \equiv 2,3 \pmod 4$ the only integral elements are $a + b\sqrt{d}$ where a, b are integers; that is, they are the elements of $\mathbb{Z}[\sqrt{d}]$. On the other hand, if $d \equiv 1 \pmod 4$, we find the additional algebraic integers $(\alpha + \beta\sqrt{d})/2$, where α and β are odd. It is easy to check that these are all elements of $\mathbb{Z} + \mathbb{Z} \cdot \frac{1+\sqrt{d}}{2}$.

6.3. **localization and integral elements.** Recall that $A \subseteq B$ is called an integral extension if every element of B is integral over A. Some easy examples are $A \subseteq A$, $\mathbb{Z} \subseteq \mathbb{Z}[\sqrt{5}]$, $\mathbb{Z} \subseteq \mathbb{Z}[\frac{1+\sqrt{5}}{2}]$, $\mathbb{C}[t] \subseteq \mathbb{C}[t,y]/(y^2 - t^3 - 1)$.

Lemma 6.3.1. (Integral extensions are preserved under localization) *Let $A \subset B$ be an integral extension and let $S \subset A$ be a multiplicative set. Then*

$$A[S^{-1}] \subset B[S^{-1}]$$

[15] One version of Gauss's lemma is: *A non-constant polynomial in $\mathbb{Z}[x]$ is irreducible in $\mathbb{Z}[x]$ if and only if it is both irreducible in $\mathbb{Q}[x]$ and primitive in $\mathbb{Z}[x]$.* To apply it we make the further assumption that $g(x)$ is monic and *irreducible* in $\mathbb{Z}[x]$, which we can surely make. It is thus irreducible in $\mathbb{Q}[x]$ and we even get that $f(x) = g(x)$ right away.

is an integral extension.

Proof. Let $\frac{b}{s} \in B[S^{-1}]$. The element b satisfies a polynomial

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in A[x].$$

One check that $\frac{b}{s}$ satisfies

$$x^n + \frac{a_{n-1}}{s}x^{n-1} + \cdots + \frac{a_0}{s^n} \in A[S^{-1}][x].$$

Thus, $\frac{b}{s}$ is integral over $A[S^{-1}]$. □

Remark 6.3.2. For $A \subseteq B$ we will call the morphism $\mathrm{Spec}(B) \to \mathrm{Spec}(A)$ **integral** if $A \subseteq B$ is an integral extension. Let $g \in A$ then the preimage of the open set $D_A(g) \subset \mathrm{Spec}(A)$ (that can be identified with $\mathrm{Spec}(A[g^{-1}])$ is the open set $D_B(g)$ (which can be identified with $\mathrm{Spec}(B[g^{-1}])$). The lemma says that the morphism

$$D_B(g) \to D_A(g)$$

is integral as well.

Lemma 6.3.3. (Integral closure is preserved under localization) *Let $A \subseteq B$ be rings and C the integral closure of A in B. Let $S \subset A$ be a multiplicative set. Then $C[S^{-1}]$ is the integral closure of $A[S^{-1}]$ in $B[S^{-1}]$.*

Proof. We already know that $A[S^{-1}] \subseteq C[S^{-1}]$ is an integral extension. Let $\frac{b}{s} \in B[S^{-1}]$ be integral over $A[S^{-1}]$; it solves some polynomial

$$x^n + \frac{a_{n-1}}{s_{n-1}}x^{n-1} + \cdots + \frac{a_0}{s_0}, \qquad a_i \in A, s_i \in S.$$

As S is multiplicative, we can pass to a common denominator and assume that all the s_i are equal and so $\frac{b}{s}$ solves

$$x^n + \frac{a_{n-1}}{s_0}x^{n-1} + \cdots + \frac{a_0}{s_0}, \qquad a_i \in A, s_0 \in S.$$

Let $\beta = s_0 b$. Then, in the ring $B[S^{-1}]$ we have

$$0 = \frac{0}{1} = \frac{b^n}{s^n} + \frac{a_{n-1}}{s_0}\frac{b^{n-1}}{s^{n-1}} + \cdots + \frac{a_0}{s_0} = \frac{1}{s^n s_0^n}(\beta^n + a'_{n-1}\beta^{n-1} + \cdots + a'_0),$$

where $a'_i = a_i s_0^{n-i-1} s^{n-i}$. This means that for some $t \in S$ we have $t(\beta^n + a'_{n-1}\beta^{n-1} + \cdots + a'_0) = 0$ in the ring B, which implies the equality in B

$$(t\beta)^n + a'_{n-1}t(t\beta)^{n-1} + \cdots + a'_0 t^n = 0.$$

Note that the coefficients $a'_{n-i}t^i$ are in A. That is, $t\beta$ is integral over A and so belongs to C. Consequently, $\frac{b}{s} = \frac{t\beta}{s_0 ts} \in C[S^{-1}]$. □

Let A be an integral domain. We say that A is **integrally closed** if A is equal to its integral closure in $\mathrm{Frac}(A)$. For example, any UFD is integrally closed.

Corollary 6.3.4. (Integrally closed is a local property) *Let A be an integral domain and $B = Frac(A)$. The following are equivalent:*

(1) A is integrally closed.

(2) $A_{\mathfrak{p}}$ is integrally closed for all prime ideals $\mathfrak{p}$.

(3) $A_{\mathfrak{m}}$ is integrally closed for all maximal ideals $\mathfrak{m}$.

Proof. Assume that A is integrally closed. Then A is the integral closure of A in B. Lemma 6.3.3 implies that $A[S^{-1}]$ is the integral closure of $A[S^{-1}]$ in $B[S^{-1}] = B$. In particular, (2) holds.

If (2) holds then also (3) because any maximal ideal is prime. Assume (3) and let C be the integral closure of A in B. Then, using the lemma again, for every maximal ideal $C_{\mathfrak{m}}$ is the integral closure of $A_{\mathfrak{m}}$ in $B_{\mathfrak{m}} = B$. Thus, $C_{\mathfrak{m}} = A_{\mathfrak{m}}$ for every maximal ideal. It follows that C/A viewed as an A-module, satisfies $(C/A)_{\mathfrak{m}} = C_{\mathfrak{m}}/A_{\mathfrak{m}} = 0$. Proposition 3.4.1 and Remark 3.4.2 then give that $C/A = 0$, namely that A is equal to its integral closure in B. □

6.4. **The going-up and going-down theorems of Cohen-Seidenberg.** Throughout this section $A \subseteq B$ is an integral extension. Our goal is to understand the morphism

$$f : \mathrm{Spec}(B) \to \mathrm{Spec}(A)$$

Otherwise said, we want to understand how prime ideals behave in integral extensions.

Proposition 6.4.1. *Assume that B is an integral domain. Then*

$$A \text{ is a field} \iff B \text{ is a field}.$$

Geometrically: if $A \subset B$ is an integral extension of integral domains, $\mathrm{Spec}(A)$ is a point if and only if $\mathrm{Spec}(B)$ is a point.

We remark that the assumption that B is an integral domain is necessary. For example, let k be a field; the extension $k \subseteq k[\epsilon]$ is integral, but $k[\epsilon]$ is not a field since $\epsilon^2 = 0$.

Proof. Suppose that A is a field and $b \in B$, $b \neq 0$. There is a minimal positive integer n such that for some $a_i \in A$,

$$b^n + a_{n-1}b^{n-1} + \cdots + a_0 = 0.$$

Then $a_0 \neq 0$, otherwise $b(b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1) = 0$ and, as B is an integral domain, $b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1$, contradicting the choice of n. Then

$$b(b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1)(-a_0)^{-1} = 1,$$

so b is invertible.

Conversely, suppose that B is a field and let $a \in A$, $a \neq 0$. Then $a^{-1} \in B$ and is integral over A. So for some $a_i \in A$

$$(a^{-1})^n + a_{n-1}(a^{-1})^{n-1} + \cdots + a_0 = 0.$$

Multiplying by a^{n-1} and rearranging gives

$$a^{-1} = -(a_{n-1} + \cdots + a_0 a^{n-1}) \in A.$$

□

Proposition 6.4.2. *Let $q \lhd B$ be a prime ideal and $\mathfrak{p} = \mathfrak{q} \cap A$. Then $\mathfrak{q}$ is maximal if and only if $\mathfrak{p}$ is maximal.*

Geometrically: If $A \subseteq B$ is an integral extension, the image of closed point is a closed point and the fiber over a closed point consists of closed points only.

Proof. We have a natural inclusion $A/\mathfrak{p} \subseteq B/\mathfrak{q}$ of integral domains. This is also an integral extension since if $b \in B$ solves $x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in A[x]$, $\bar{b}$ (the image of b in $B/\mathfrak{q}$) solves $x^n + \bar{a}_{n-1}x^{n-1} + \cdots + \bar{a}_0 \in (A/\mathfrak{p})[x]$. We can now use the previous proposition, recalling that an ideal is maximal if and only if the quotient is a field.

The geometric interpretation follows once we note that $[\mathfrak{p}] = f([\mathfrak{q}])$. □

Proposition 6.4.3. *Let $\mathfrak{q}_1 \subseteq \mathfrak{q}_2$ be prime ideals of B such that $\mathfrak{q}_1 \cap A = \mathfrak{q}_2 \cap A$ then $\mathfrak{q}_1 = \mathfrak{q}_2$.*

Geometrically: closure in B happens horizontally relative to A.

Proof. Let $\mathfrak{p} = \mathfrak{q}_1 \cap A = \mathfrak{q}_2 \cap A$. If $\mathfrak{p}$ were maximal then we'd be done by the previous proposition because both $\mathfrak{q}_1, \mathfrak{q}_2$ would have been maximal and, as $\mathfrak{q}_1 \subseteq \mathfrak{q}_2$, we would get $\mathfrak{q}_1 = \mathfrak{q}_2$. But $\mathfrak{p}$ doesn't have to be maximal, so we are going to make it maximal by localization.

Let $S = A \setminus \mathfrak{p}$, which is a multiplicative set both in A and in B. By Lemma 6.3.1, $A[S^{-1}] \subseteq B[S^{-1}]$ is an integral extension. In addition $\mathfrak{p}[S^{-1}]$ is the maximal ideal of $A[S^{-1}]$ and $\mathfrak{q}_i[S^{-1}]$ are prime ideals of $B[S^{-1}]$ such that $\mathfrak{q}_i[S^{-1}] \cap A[S^{-1}] = \mathfrak{p}[S^{-1}]$.[16] It follows now that both $\mathfrak{q}_i[S^{-1}]$ are maximal ideals of $B[S^{-1}]$ and $\mathfrak{q}_1[S^{-1}] \subseteq \mathfrak{q}_2[S^{-1}]$. Thus, $\mathfrak{q}_1[S^{-1}] = \mathfrak{q}_2[S^{-1}]$. As both are prime ideals and both q_i are disjoint with S we get $\mathfrak{q}_1 = \mathfrak{q}_2$.

The geometric interpretation follows once we note that $\overline{[\mathfrak{q}_1]} = V(\mathfrak{q}_1) = \{q_2 : q_2 \supseteq q_1\}$. So no point in the closure of $\mathfrak{q}_1$, except $\mathfrak{q}_1$ itself, can lie in the fiber over $f(\mathfrak{q}_1)$. □

Corollary 6.4.4. *Let $\mathfrak{q}_1 \subsetneq \mathfrak{q}_2 \subsetneq \cdots \subsetneq \mathfrak{q}_n$ be prime ideals of B. Let $\mathfrak{p}_i = \mathfrak{q}_i \cap A$ then $\mathfrak{p}_1 \subsetneq \mathfrak{p}_2 \subsetneq \cdots \subsetneq \mathfrak{p}_n$.*

Proposition 6.4.5. *Let $\mathfrak{p}$ be a prime ideal of A. There is a prime ideal $\mathfrak{q}$ of B such that $\mathfrak{q} \cap A = \mathfrak{p}$.*

Geometrically: The morphism $f : \mathrm{Spec}(B) \to \mathrm{Spec}(A)$ is surjective.

Proof. Note that if $\mathfrak{p}$ were a maximal ideal of A we could have taken $\mathfrak{q}$ to be a maximal ideal of B containing $\mathfrak{p}$. Then $\mathfrak{q} \cap A$ is a (prime) ideal containing $\mathfrak{p}$ and so must be equal to $\mathfrak{p}$. However, there is a problem with this argument as *a priori* we don't know that any such ideal $\mathfrak{q}$ exists; it's existence requires that $A \subseteq B$ is an integral extension (for example, the statement that $\mathfrak{q}$ exists would fail for $\mathbb{Z} \subset \mathbb{Q}$) and is a consequence of the proof we give below. The other issue is that $\mathfrak{p}$ need not be maximal; we shall make $\mathfrak{p}$ maximal by localizing. After these motivating comments we start the proof proper:

Let $S = A \setminus \mathfrak{p}$. $\mathfrak{p}[S^{-1}]$ is the unique maximal ideal of $A[S^{-1}]$ and we can take $\mathfrak{n}$ to be any maximal ideal of $B[S^{-1}]$ containing $\mathfrak{p}[S^{-1}]$ and so $\mathfrak{n} \cap A[S^{-1}] = \mathfrak{p}[S^{-1}]$. Such an $\mathfrak{n}$ exists; in fact, if $\mathfrak{n}$ is any maximal ideal of $B[S^{-1}]$ then $\mathfrak{n} \cap A[S^{-1}]$ is maximal ideal because $A[S^{-1}] \subset B[S^{-1}]$ is an integral extension. Thus, necessarily

$$\mathfrak{n} \cap A[S^{-1}] = \mathfrak{p}[S^{-1}].$$

Let $\mathfrak{q} = \mathfrak{n}^c$, a prime ideal of B disjoint from S, such that $\mathfrak{q}^e = \mathfrak{q}[S^{-1}] = \mathfrak{n}$. Then

$$(\mathfrak{q} \cap A)[S^{-1}] = \mathfrak{q}[S^{-1}] \cap A[S^{-1}] = \mathfrak{p}[S^{-1}].$$

[16] This follows from a general property: *if $M_1, M_2 \subset M$ are A-modules and S is a multiplicative set of A then*

$$(M_1 \cap M_2)[S^{-1}] = M_1[S^{-1}] \cap M_2[S^{-1}]$$

inside $M[S^{-1}]$. To see that note that the inclusion $\subseteq$ is immediate. So let $\frac{m}{s} \in M_1[S^{-1}] \cap M_2[S^{-1}]$, say $\frac{m}{s} = \frac{m_1}{s_1} = \frac{m_2}{s_2}$ for $m_i \in M_i, s_i \in S$. Then for some $t \in S$, $ts_2m_1 = ts_1m_2$ and this element is in $M_1 \cap M_2$. As $\frac{m}{s} = \frac{ts_2m_1}{ts_2s_1} \in (M_1 \cap M_2)[S^{-1}]$, we get the other inclusion. Now use this property for the A-modules $A = M_1, \mathfrak{q}_i = M_2, B = M$.

As $\mathfrak{q} \cap A$ is a prime ideal of A disjoint from S, we must have

$$\mathfrak{q} \cap A = \mathfrak{p}.$$

□

Theorem 6.4.6. (Going-up theorem of Cohen-Seidenberg) *Let $A \subseteq B$ be an integral extension. Let $\mathfrak{p}_1 \subseteq \mathfrak{p}_2$ be prime ideals of A and let $\mathfrak{q}_1$ be a prime ideal of B such that $\mathfrak{q}_1 \cap A = \mathfrak{p}_1$. Then, there is a prime ideal $\mathfrak{q}_2$ of B such that $\mathfrak{q}_1 \subseteq \mathfrak{q}_2$ and $\mathfrak{q}_2 \cap A = \mathfrak{p}_2$. Namely, we can fill in the ? in the following diagram:*

$$\begin{array}{ccccc} B & & \mathfrak{q}_1 & \subseteq & ? \\ & & | & & | \\ A & & \mathfrak{p}_1 & \subseteq & \mathfrak{p}_2 \end{array}$$

Geometrically: if $f([q_1]) = [\mathfrak{p}_1]$ then $f(V(\mathfrak{q}_1)) = V(\mathfrak{p}_1)$.[17]

Proof. The extension $A/\mathfrak{p}_1 \subseteq B/\mathfrak{q}_1$ is an integral extension and $\mathfrak{p}_2/\mathfrak{p}_1$ is a prime ideal of $A/\mathfrak{p}_1$. By Proposition 6.4.5, there is a prime ideal of $B/\mathfrak{q}_1$, that we can always write as $\mathfrak{q}_2/\mathfrak{q}_1$ where $\mathfrak{q}_2$ is a prime ideal of B, such that

$$\mathfrak{q}_2/\mathfrak{q}_1 \cap A/\mathfrak{p}_1 = \mathfrak{p}_2/\mathfrak{p}_1.$$

This implies that $\mathfrak{q}_2 \cap A = \mathfrak{p}_2$. To see that, note that the meaning of having an *inclusion* $A/\mathfrak{p}_1 \subseteq B/\mathfrak{q}_1$ is that in $B/\mathfrak{q}_1$ we have $A/\mathfrak{p}_1 = (A + \mathfrak{q}_1)/\mathfrak{q}_1$. And similarly, $\mathfrak{p}_2/\mathfrak{p}_1 = (\mathfrak{p}_2 + \mathfrak{q}_1)/\mathfrak{q}_1$. So we can say that

$$\mathfrak{q}_2/\mathfrak{q}_1 \cap (A + \mathfrak{q}_1)/\mathfrak{q}_1 = (\mathfrak{p}_2 + \mathfrak{q}_1)/\mathfrak{q}_1.$$

This implies that $\mathfrak{q}_2 \cap (A + \mathfrak{q}_1) = \mathfrak{p}_2 + \mathfrak{q}_1$. Let then $a \in \mathfrak{q}_2 \cap A$. Then $a \in \mathfrak{q}_2 \cap (A + \mathfrak{q}_1)$ and so $a = b + c$ for some $b \in \mathfrak{p}_2, c \in \mathfrak{q}_1$. But this implies that $c \in A$ too and so $c \in A \cap \mathfrak{q}_1 = \mathfrak{p}_1$ and thus $a \in \mathfrak{p}_2 + \mathfrak{p}_1 = \mathfrak{p}_2$. Conversely, if $a \in \mathfrak{p}_2$ then $a \in \mathfrak{p}_2 + \mathfrak{q}_1$ and so $a \in \mathfrak{q}_2 \cap (A + \mathfrak{q}_1)$ and a fortiori $a \in \mathfrak{q}_2 \cap A$. We proved that

$$\mathfrak{q}_2 \cap A = \mathfrak{p}_2.$$

□

Corollary 6.4.7. *Let $A \subseteq B$ be an integral extension. Let $\mathfrak{p}_1 \subset \mathfrak{p}_2 \subset \cdots \subset \mathfrak{p}_n$ be a chain of prime ideals of A and let $\mathfrak{q}_1 \subset \mathfrak{q}_2 \subset \cdots \subset \mathfrak{q}_m$ be a chain of prime ideals of B, where $m \leq n$, such that*

$$\mathfrak{q}_i \cap A = \mathfrak{p}_i, \quad 1 \leq i \leq m.$$

Then the chain can be extended to $\mathfrak{q}_1 \subset \mathfrak{q}_2 \subset \cdots \subset \mathfrak{q}_n$ so that $\mathfrak{q}_i \cap A = \mathfrak{p}_i, \quad 1 \leq i \leq n$.

Proposition 6.4.8. *Let $A \subset B$ be an integral extension then the morphism $\mathrm{Spec}(B) \to \mathrm{Spec}(A)$ is a closed map.*

[17] It is a priori clear that $f(V(\mathfrak{q}_1)) \subset V(\mathfrak{p}_1)$ and contains $[\mathfrak{p}_1]$. If we knew that f is a closed map, it would follow that $f(V(\mathfrak{q}_1)) = V(\mathfrak{p}_1)$. In the other direction, we only have the statement that f closed on sets of the form $V(\mathfrak{q})$ where $\mathfrak{q}$ is a prime ideal. To be precise, $f(V(\mathfrak{q})) = V(\mathfrak{q} \cap A)$. But the general closed set is of the form $V(J)$, where J is an ideal of B that need not be prime. Thus, we didn't quite prove that f is a closed map. This is in fact true but requires extra work. One way to proceed is given an ideal $J \lhd B$, let $I \lhd A$ and consider the extension $A/I \hookrightarrow B/J$.

Proof. Every closed set of B is of the form $V(\mathfrak{b})$ for some ideal $\mathfrak{b}\lhd B$. If $\varphi : \mathrm{Spec}(B) \to \mathrm{Spec}(A)$ is the map corresponding to the inclusion $A \hookrightarrow B$, then

$$\varphi(V(\mathfrak{b})) = \{\mathfrak{q} \cap A : \mathfrak{q}\lhd B \text{ prime}, \mathfrak{q} \supseteq \mathfrak{b}\} \subseteq \{\mathfrak{p}\lhd A \text{ prime}, \mathfrak{p} \supset \mathfrak{b} \cap A\} = V(\mathfrak{a}),$$

where $\mathfrak{a} = \mathfrak{b} \cap A$. So, we need to show that every prime ideal $\mathfrak{p}$ of A that contains $\mathfrak{a}$ is of the form $\mathfrak{q} \cap A$ where $\mathfrak{q} \supseteq \mathfrak{b}$ is a prime ideal. Note that the inclusion

$$A/\mathfrak{a} \hookrightarrow B/\mathfrak{b},$$

is an integral extension. Thus, $\mathrm{Spec}(B/\mathfrak{b}) \to \mathrm{Spec}(A/\mathfrak{a})$ is surjective and there is a prime ideal $\bar{\mathfrak{q}}\lhd B/\mathfrak{b}$ such that $\bar{\mathfrak{q}} \cap A/\mathfrak{a} = \bar{\mathfrak{p}}$, where $\bar{\mathfrak{p}}$ is the image of $\mathfrak{p}$ in $A/\mathfrak{a}$ (which is a prime ideal). Let $\mathfrak{q}\lhd B$ be the preimage of $\bar{\mathfrak{q}}$ in B, which is a prime ideal. Furthermore, $\mathfrak{q} \cap A = \mathfrak{p}$, as can be verified by the same reasoning as in Theorem 6.4.6. Thus,

$$\varphi(V(\mathfrak{b})) = V(\mathfrak{b} \cap A),$$

and in particular φ is a closed map. □

The going-down theorem of Cohen-Seidenberg is much more difficult to prove. We state it only for completeness. The proof can be found in Atiyah-MacDonald.

Theorem 6.4.9. (Going-down theorem of Cohen-Seidenberg) *Let $A \subseteq B$ be an integral extension of integral domains such that A is integrally closed. Let $\mathfrak{p}_1 \subseteq \mathfrak{p}_2$ be prime ideals of A. Let $\mathfrak{q}_2$ be a prime ideal of B such that $\mathfrak{q}_2 \cap A = \mathfrak{p}_2$. Then we can find a prime ideal $\mathfrak{q}_1 \subseteq \mathfrak{q}_2$ of B such that $\mathfrak{q}_1 \cap A = \mathfrak{p}_1$. Namely, we can fill in the ? in the diagram:*

$$\begin{array}{ccccc} B & & ? & \subseteq & \mathfrak{q}_2 \\ & & | & & | \\ A & & \mathfrak{p}_1 & \subseteq & \mathfrak{p}_2 \end{array}$$

Geometrically: If a point $y \in \mathrm{Spec}(A)$ is in the closure of a point x, then any point in the fibre of f over y is in the closure of a point in the fiber of f over x.

7. Noetherian rings

7.1. **Noetherian rings and modules.**

7.1.1. *Noetherian rings.* This theory can be developed for non-commutative rings. One talks then about left Noetherian rings and right Noetherian rings, but as our main application is to commutative rings, we restrict our discussion to this setting.

Let R be a commutative ring. R is called **noetherian** if any increasing chain of ideals of R stabilizes. Namely, if

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$$

are ideals of R then there is some n such that

$$I_n = I_{n+1} = I_{n+2} = \dots.$$

Example 7.1.1. $\mathbb{Z}$ is a noetherian ring. $\mathbb{C}[x]$ is a noetherian ring. In fact, any PID R is a noetherian ring because in that case the ideal $I = \cup_{a=1}^{\infty} I_a$ is principal and therefore equal to Rf for some $f \in R$. But then $f \in I_n$ for some n and it follows that that $I_n = I$. That implies that $I_n = I_m, \forall m \geq n$.

Proposition 7.1.2. *R is noetherian if and only if every ideal is finitely generated.*

Proof. Suppose that every ideal is finitely generated and consider a chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$. Let $I_\infty = \cup_{a=1}^{\infty} I_a$; it is an ideal of R and so $I = \langle f_1, \dots, f_n \rangle$ for some f_i. Each f_i belongs to some ideal I_{n_i} and so there is an N such that $f_i \in I_N, i = 1, \dots, n$. It follows that $I_N = I_{N+1} = \dots = I_\infty$.

Conversely, suppose that R is noetherian and let I be an ideal of R. Define a sequence of ideals inductively. $I_0 = \{0\}$. If there is an element f_1 of I that is not in I_0, let $I_1 = \langle f_1 \rangle$. Inductively, suppose that we defined $I_a = \langle f_1, \dots, f_a \rangle \subseteq I$, $0 \leq a \leq n$, a strictly increasing sequence of ideals. If there is an element $f_{n+1} \in I \setminus I_n$, let $I_{n+1} = \langle f_1, \dots, f_n, f_{n+1} \rangle \subseteq I$. Then we have the sequence of ideals $I_0 \subsetneqq I_1 \subsetneqq I_2 \subsetneqq \dots \subsetneqq I_{n+1} \subseteq I$. The noetherian hypothesis implies this process must stop at some N. It must be then that $I_N = \langle f_1, \dots, f_N \rangle = I$ and so I is finitely generated. □

Proposition 7.1.3. *If R is noetherian so is R/I for any ideal $I \lhd R$.*

Proof. A sequence of ideals of R/I can be lifted to sequence of ideals of R that must stabilize at some point. If follows that the original sequence stabilizes too. □

Example 7.1.4. An example of a ring that is not noetherian is the ring $\mathbb{C}[t^\alpha : \alpha \in \mathbb{Q}^+]$. The ideal $\langle \{t^\alpha : \alpha \in \mathbb{Q}^+\} \rangle$ is not finitely generated. Note that this ring is the direct limit of noetherian rings

$$\mathbb{C}[t] \hookrightarrow \mathbb{C}[t^{1/2}] \hookrightarrow \mathbb{C}[t^{1/6}] \hookrightarrow \mathbb{C}[t^{1/24}] \dots$$

(each $\mathbb{C}[t^{1/n!}] \cong \mathbb{C}[x]$ and so is noetherian).

For another example one can take the ring of complex polynomials in infinitely many variables $\mathbb{C}[x_1, x_2, \dots] = \varinjlim_n \mathbb{C}[x_1, \dots, x_n]$. Hilbert's basis theorem says that $\mathbb{C}[x_1, \dots, x_n]$ is noetherian, but the limit is not. The ideal $\langle x_1, x_2, \dots \rangle$ is not finitely generated.

7.1.2. *Noetherian modules.* Let R be a commutative ring. An R-module M is called **noetherian** if every sub R-module $M' \subseteq M$ is a finitely generated R-module. Equivalently (compare Proposition 7.1.2), M is noetherian if every increasing chain of submodules of M, $M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$, stabilizes. In particular, R is a noetherian ring if and only if R is a noetherian R-module.

Lemma 7.1.5. *Let $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ be an exact sequence of R-modules. Then M_2 is noetherian if and only if both M_1 and M_3 are noetherian.*

Proof. We will treat f as an inclusion map and so M_1 is a submodule of M_2.

Suppose M_2 is noetherian. An increasing sequence of submodules of M_1 can be viewed as a sequence of submodules of M_2, hence it stabilizes. An incereasing sequence of submodules of M_3 can be lifted to an increasing sequence of submodules of M_2 that must stabilize. Hence the sequence in M_3 stabilizes too. That is, if M_2 is noetherian so are M_1 and M_3.

Now suppose that M_1 and M_3 are noetherian and let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ be an increasing sequence of submodules of M_2. Then

$$A_1 \cap M_1 \subseteq A_2 \cap M_1 \subseteq A_3 \cap M_1 \subseteq \cdots$$

stabilizes. Say, for all $n \geq N$ we have $A_n \cap M_1 = A_N \cap M_1$. At the same time, the sequence $g(A_1) \subseteq g(A_2) \subseteq g(A_3) \subseteq \cdots$ in M_3 also stabilizes. So, increasing N if necessary, we may assume that for all $n \geq N$ we have $g(A_n) = g(A_N)$. We claim that this implies that

$$A_n = A_N, \quad \forall n \geq N.$$

The inclusion $A_n \supseteq A_N$ is given. To show the opposite inclusion, let $a \in A_n$ then $g(a) \in g(A_N)$ and so there is some element $b \in A_N$ such that $g(a) = g(b)$. Then $g(a-b) = 0$; namely, $a - b \in M_1 \cap A_n$. But $M_1 \cap A_n = M_1 \cap A_N$ and so $a - b = c$ for some $c \in A_N$. It now follows that $a = b + c \in A_N$. □

Corollary 7.1.6. *Let R be a noetherian ring then for every $N \in \mathbb{N}$, R^N is a noetherian R-module.*

Proof. Argue by induction; the cases $N = 0, 1$ are clear. Assume true for some $N \geq 1$. Then R^{N+1} sits in an exact sequence

$$0 \to R^N \to R^{N+1} \to R \to 0,$$

where the maps are $(r_1, \ldots, r_n) \mapsto (r_1, \ldots, r_n, 0)$ and $(r_1, \ldots, r_{n+1}) \mapsto r_{n+1}$. As R and R^N are noetherian, so it R^{N+1}. □

Corollary 7.1.7. *Let R be a noetherian ring then any finitely generated R-module is noetherian.*

Proof. Let M be a finitely generated R-module. Then there is a surjective homomorphism of R-modules $R^N \to M$ for some N. As R^N is noetherian, so is M by Lemma 7.1.5. □

It follows that if R is noetherian then a submodule N of a finitely generated module M is finitely generated. Indeed, M is noetherian, therefore so is N. If R is not noetherian this may fail. For example, take $R = \mathbb{C}[x_1, x_2, \ldots]$, which is a finitely generated module over itself (it is generated by 1), but its R submodule $(x_1, x_2, \ldots)$ is not finitely generated.

7.2. **Hilbert's basis theorem.** Hilbert's basis theorem was a landmark result in algebra. it implies that any ideal in a polynomial ring $k[x_1, \ldots, x_n]$, where k is a field, is finitely generated. Such an ideal may arise as the set of all relations satisfied by the polynomials in $k[x_1, \ldots, x_n]$ when restricted to a subvariety, or even just a subset, of an affine space. The finite generation then implies that there is always a *finite* list of relations such that all other relations are consequence of which.

At the time Hilbert proved his basis theorem such problems where at the forefront of research in algebra and algebraic geometry and so Hilbert's result was revolutionary. It also attracted much opposition, because the proof gave no indication of how such a finite list of generators may be found.[18] Since then the situation has changed; the introduction of Gröbner bases allowed making such tasks as finding a list of generators effectively computable.

Theorem 7.2.1. (Hilbert's basis theorem) *Let R be a commutative noetherian ring then the ring of polynomials $R[x]$ is also noetherian.*

[18](Apparently) when P. Gordan, one of the leading algebraists of the time, first saw Hilbert's proof, he said, "This is not Mathematics, but theology!" On the other hand, Gordan said, in 1899 when he published a simplified proof of Hilbert's Theorem, "I have convinced myself that theology also has its advantages."

Proof. Let $\mathfrak{a} \lhd R[x]$. We need to show that $\mathfrak{a}$ is finitely generated. Let I be the collection of all leading coefficients of polynomials in $\mathfrak{a}$. Then, in fact, I is an ideal: Suppose that $a_i \in \mathfrak{a}$, $i = 1,2$ then we have some polynomials

$$f_1(x) = a_1 x^{n_1} + l.o.t. \in \mathfrak{a}, \qquad f_2(x) = a_2 x^{n_2} + l.o.t. \in \mathfrak{a}.$$

Without loss of generality $n_1 \geq n_2$. Then also, for any $r \in R$,

$$rf_1(x) = ra_1 x^{n_1} + l.o.t. \in \mathfrak{a}, \qquad f_1(x) + x^{n_1 - n_2} f_2(x) = (a_1 + a_2)x^{n_1} + l.o.t \in \mathfrak{a}.$$

This implies that $ra_1 \in I, a_1 + a_2 \in I$ and we see that I is an ideal.

Being an ideal of R, I is finitely generated. Say

$$I = \langle a_1, \ldots, a_n \rangle.$$

Then for any $1 \leq i \leq n$

$$\exists\, f_i(x) = a_i x^{r_i} + l.o.t. \in \mathfrak{a}.$$

Let $r = \max\{r_1, \ldots, r_n\}$ and let

$$\mathfrak{a}' = \langle f_1, \ldots, f_n \rangle \subseteq \mathfrak{a}.$$

Let $f = ax^m + l.o.t.$ be an element of $\mathfrak{a}$. Then $a \in I, a = \sum_{i=1}^n u_i a_i$ for some $u_i \in R$. If $m \geq r$,

$$f - \sum u_i x^{m - r_i} f_i \in \mathfrak{a}$$

and has degree less than m. Thus, after repeating this process, we may write

$$f = g + h, \quad h \in \mathfrak{a}', g \in \mathfrak{a}, \deg(g) < r.$$

Let

$$M = R \oplus Rx \oplus \cdots \oplus Rx^{r-1}.$$

Since M is a finitely generated R-module, M is a noetherian R-module. Therefore, $M \cap \mathfrak{a}$, which is an R-submodule of M, is also noetherian and is finitely generated as an R-module by, say, $g_1, \ldots, g_m$. We have shown that

$$\mathfrak{a} = \mathfrak{a}' + M \cap \mathfrak{a} = \langle f_1, \ldots, f_n, g_1, \ldots, g_m \rangle.$$

□

Corollary 7.2.2. *Let k be a field and let $x_1, \ldots, x_n$ be n variables. The ring $k[x_1, \ldots, x_n]$ is noetherian. Moreover, let $I \lhd k[x_1, \ldots, x_n]$ be an ideal then the ring $k[x_1, \ldots, x_n]/I$ is noetherian.*

Proof. The field k is noetherian as its only ideals are $\{0\}$ and k. By Hilbert's basis theorem so is $k[x_1]$ and therefore so is $k[x_1][x_2] = k[x_1, x_2]$ and so on. The second statement follows from Proposition 7.1.3. □

Corollary 7.2.3. *Let $R \subseteq S$ be commutative rings. Assume that R is noetherian and S is a finitely generated R-algerba, then S is noetherian.*

Proof. S is a quotient of the polynomial ring $R[x_1, \ldots, x_n]$ for some n. □

7.3. **Noether's normalization lemma.** The following lemma was proven by Emmy Noether.

Theorem 7.3.1. (Noether's normalization lemma) *Let k be a field and let A be a finitely generated k-algebra. There exists $y_1,\dots,y_r \in A$, algebraically independent over k, such that A is integral over $k[y_1,\dots,y_r]$.*

Geometrically: $\mathrm{Spec}(A)$ *affords a (surjective) integral morphism to an affine space* $\mathbb{A}^r_k := \mathrm{Spec}(k[y_1,\dots,y_r])$.[19]

We prove Noether's Lemma for the case k is an infinite field. It is true for finite fields too, but the proof is a bit different. Before the proof we introduce some notation. If $t_1,\dots,t_n$ are variables, $I=(i_1,\dots,i_n), i_j \in \mathbb{N}$ a multi-index, we let

$$|I| = i_1+\cdots+i_n, \quad t^I = t_1^{i_1}\cdots t_n^{i_n}.$$

With this notation any $f \in k[t_1,\dots,t_n]$ can be written as

$$f = \sum_I a_I t^I = \sum_{|I|=e} a_I t^I + \sum_{|I|<e} a_I t^I,$$

where we implicitly assume that $a_I \neq 0$ for some I with $|I|=e$, and we define then

$$F = \sum_{|I|=e} a_I t^I.$$

Proof. Assume k is an infinite field. Write

$$A = k[x_1,\dots,x_n]$$

for some *generators* x_i – they need not be algebraically independent. Without loss of generality, $x_1,\dots,x_r$ are algebraically independent and $x_{r+1},\dots,x_n$ are algebraic over $k[x_1,\dots,x_r]$. We proceed by induction on n.

If $n=r$, there is nothing to prove. Suppose $n>r$. Since x_n is algebraic over $k[x_1,\dots,x_{n-1}]$ there exists some $f \in k[t_1,\dots,t_n]$ such that $f(x_1,\dots,x_n)=0$ and $\deg_{t_n}(f) \neq 0$. Since k is infinite, there exist some $\lambda_1,\dots,\lambda_{n-1}$ such that

$$F(\lambda_1,\dots,\lambda_{n-1},1) \neq 0.$$

This requires proof, but we defer it to the end. Let

$$x_1' = x_1 - \lambda_1 x_n, \;\; x_2' = x_2 - \lambda_2 x_n, \;\; \dots, \;\; x_{n-1}' = x_{n-1} - \lambda_{n-1} x_n.$$

Then

$$\begin{aligned}
0 &= f(x_1'+\lambda_1 x_n, \dots, x_{n-1}'+\lambda_{n-1}x_n, x_n) \\
&= \sum_{|I|=e} a_I (x_1'+\lambda_1 x_n)^{i_1}\cdots(x_{n-1}'+\lambda_{n-1}x_n)^{i_{n-1}} x_n^{i_n} + \sum_{|I|<e} a_I x_n^{i_n} \prod_{a=1}^{n-1} (x_a'+\lambda_a x_n)^{i_a} \\
&= x_n^e \sum_{|I|=e} a_I \lambda_1^{i_1}\cdots\lambda_{n-1}^{i_{n-1}} 1^{i_n} + \text{l.o.t. in } x_n \text{ with coefficients in } k[x_1',\dots,x_{n-1}'] \\
&= x_n^e F(\lambda_1,\dots,\lambda_{n-1},1) + \text{l.o.t. in } x_n \text{ with coefficients in } k[x_1',\dots,x_{n-1}'].
\end{aligned}$$

As $F(\lambda_1,\dots,\lambda_{n-1},1)\neq 0$ we can divide by it; we conclude that there is a relation of the form

$$0 = x_n^e + \alpha_{e-1}x_n^{e-1} + \cdots + \alpha_0, \qquad \alpha_i \in k[x_1',\dots,x_{n-1}'].$$

[19] In this case, using that the rings involved are noetherian, one can also prove that such a morphism has finite fibers. Thus, every affine noetherian scheme over k, is a finite cover of an affine space by a surjective morphism with finite fibers (and other nice properties) that is typically ramified.

It follows that x_n, and so A, is integral over the subring $k[x'_1,\dots,x'_{n-1}]$. By induction, there exist $y_1,\dots,y_r \in k[x'_1,\dots,x'_{n-1}]$ such that the y_i are algebraically independent over k and such that $k[x'_1,\dots,x'_{n-1}]$ is integral over $k[y_1,\dots,y_r]$. But then A is integral over $k[y_1,\dots,y_r]$.

All that remains is to prove the following lemma:

Lemma 7.3.2. *Let k be an infinite field, $F(t_1,\dots,t_{a+1})$ a non-zero homogeneous polynomial over k. There exist $\lambda_1,\dots,\lambda_a \in k$ such that*

$$F(\lambda_1,\dots,\lambda_a,1) \neq 0.$$

Proof. (Lemma) Write

$$F(t_1,\dots,t_{a+1}) = t_{a+1}^e \cdot f_0 + t_{a+1}^{e-1} f_1 + \cdots + f_e,$$

where $f_i \in k[t_1,\dots,t_a]$ are homogeneous polynomials of degree i. Thus,

$$f(t_1,\dots,t_a) := F(t_1,\dots,t_a,1) = f_0 + f_1 + \cdots + f_e \neq 0.$$

So all that remains is to prove that if $f(t_1,\dots,t_a) \in k[t_1,\dots,t_a]$ is a non-zero polynomial then for some $\lambda_1,\dots,\lambda_a \in k$ also $f(\lambda_1,\dots,\lambda_a) \neq 0$. We prove by induction on a.

The case $a = 0$ is clear: f is just a non-zero scalar and there's nothing to prove. If $a = 1$, the number of roots of f is at most its degree, so the assertion is also clear because k is infinite.

Suppose $a > 1$ and suppose that for all $\lambda_1,\dots,\lambda_a$ we have $f(\lambda_1,\dots,\lambda_a) = 0$. Write

$$f = t_a^d G_d(t_1,\dots,t_{a-1}) + t_a^{d-1} G_{d-1}(t_1,\dots,t_{a-1}) + \cdots + G_0(t_1,\dots,t_{a-1}).$$

For any fixed $\lambda_1,\dots,\lambda_{a-1}$ the polynomial

$$f = t_a^d G_d(\lambda_1,\dots,\lambda_{a-1}) + t_a^{d-1} G_{d-1}(\lambda_1,\dots,\lambda_{a-1}) + \cdots + G_0(\lambda_1,\dots,\lambda_{a-1}) \equiv 0,$$

(meaning, is equal to 0 for any substitution $t_a = \lambda_a$) hence it is the zero polynomial. Thus, $G_i(\lambda_1,\dots,\lambda_{a-1}) \equiv 0$ for all i and all choices of $\lambda_1,\dots,\lambda_{a-1}$. By induction, $G_i(\lambda_1,\dots,\lambda_{a-1})$ *is* the zero polynomial for every i. It follows that $f = 0$. Contradiction. □

□

Example 7.3.3. Let $A = k[x,y]/(xy-1)$. A is not integral extension of $k[x]$. Geometrically, the projection of the hyperbola on the line is not surjective; it misses $\{0\}$. Algebraically, (x) is a prime ideal of $k[x]$ for which there is no prime ideal $\mathfrak{q}$ of A such that $\mathfrak{q} \cap k[x] = (x)$, because x is a unit in A and therefore cannot belong to any prime ideal.

So how does Noether's lemma work in this example? The polynomial f appearing in the proof can be taken to be $f(x,y) = xy - 1$ and $F = xy$. Note that $F(\lambda,1) \neq 0$ for any $\lambda \neq 0$. Then

$$f(x,y) = f((x-\lambda y)+\lambda y, y) = ((x-\lambda y)+\lambda y)y - 1 = \lambda y^2 + (x-\lambda y)y - 1.$$

Then y is integral over $k[x']$ where $x' = x - \lambda y$, as y solves the polynomial in the variable t, $t^2 + \frac{x'}{\lambda}t - \frac{1}{\lambda} \in k[x'][t]$. To be concrete, take $\lambda = 1$. Then we find that the extension

$$k[x-y] \subseteq k[x,y]/(xy-1)$$

is an integral extension. We can view this inclusion as resulting from the homomorphism of rings,

$$k[t] \to k[x,y]/(xy-1), \quad t \mapsto x - y,$$

that corresponds to the geometric map from the algebraic set in $Z \subseteq k^2$, $Z = \{(x,y) : xy - 1 = 0\}$ to $\mathbb{A}^1$:

$$Z \to \mathbb{A}^1, \qquad (x,y) \mapsto x - y.$$

7.4. **Hilbert's nullstellensatz and affine space.** The following result, called "weak nullstellensatz" is a corollary of Noether's normalization lemma.

Theorem 7.4.1. (Hilbert's weak nullstellensatz) *Let k be an algebraically closed field. A maximal ideal of the polynomial ring $k[x_1,\dots,x_n]$ has the form*

$$\langle x_1-\alpha_1,\dots,x_n-\alpha_n\rangle.$$

And, conversely, every such ideal is maximal.

Proof. The last statement is easy since that ideal is the kernel of the surjective ring homomorphism

$$k[x_1,\dots,x_n]\to k,\quad f(x_1,\dots,x_n)\mapsto f(\alpha_1,\dots,\alpha_n).$$

(If this is not clear, change variables so that $\alpha_1=\cdots=\alpha_n=0$. The statement is more self-evident in this case.)

Let $\mathfrak{m}$ be a maximal ideal and let $A=k[x_1,\dots,x_n]/\mathfrak{m}$. By Noether's normalization $\mathrm{Spec}(A)$ is integral over some affine space $\mathbb{A}^N_k$. If $N>0$, $\mathbb{A}^N_k$ has many points but $\mathrm{Spec}(A)=\{*\}$ and the morphism $\mathrm{Spec}(A)\to\mathbb{A}^N_k$ cannot be surjective, hence cannot be integral. Therefore, $N=0$. That is, A is a field, integral over k. As k is algebraically closed, this implies that $A=k$.

Since $A=k$, each $x_i\equiv\alpha_i \pmod{\mathfrak{m}}$ for some $\alpha_i\in k$. That is, $x_i-\alpha_i\in\mathfrak{m}$. This implies that

$$\mathfrak{m}\supseteq\langle x_1-\alpha_1,\dots,x_n-\alpha_n\rangle,$$

which is itself a maximal ideal. It follows that $\mathfrak{m}=\langle x_1-\alpha_1,\dots,x_n-\alpha_n\rangle$. □

7.4.1. *The classical affine space of k.* Let k be an algebraically closed field. Classically, one constructs an algebraic variety $A^n_k=k^n$, the affine space over k, in the following way. The points of this space are the n-tuples k^n. The algebraic sets in this space are constructed as follows: for any radical ideal $\mathfrak{a}$ of $k[x_1,\dots,x_n]$ define $Z(\mathfrak{a})\subset k^n$,

$$Z(\mathfrak{a})=\{(\lambda_1,\dots,\lambda_n): f(\lambda_1,\dots,\lambda_n)=0,\forall f\in\mathfrak{a}\}.$$

Namely, $Z(\mathfrak{a})$ are the common zeros of all the polynomials in $\mathfrak{a}$. Conversely, given any subset $S\subset k^n$ we can construct

$$I(S)=\{f\in k[x_1,\dots,x_n]: f(s)=0,\forall s\in S\}.$$

Namely, $I(S)$ is the ideal comprised all polynomials vanishing identically on S. It is easy to see that this is a radical ideal. Hilbert's nullstellensatz says that $\mathfrak{a}\mapsto Z(\mathfrak{a})$ is a bijection between radical ideals and algebraic sets in A^n_k.

Theorem 7.4.2. (Hilbert's nullstellensatz) *Let $\mathfrak{a}$ be a radical ideal of $k[x_1,\dots,x_n]$. Then,*

$$I(Z(\mathfrak{a}))=\mathfrak{a}.$$

Proof. It is clear from the definitions that $I(Z(\mathfrak{a}))\supseteq\mathfrak{a}$. Suppose that $f\notin\mathfrak{a}$ then, by Lemma 5.2.2 there exists a prime ideal $\mathfrak{p}\supseteq\mathfrak{a}$ such that $f\notin\mathfrak{p}$.

Let $\bar f$ denote the image of f in $B=k[x_1,\dots,x_n]/\mathfrak{p}$. It is a non-zero element of this integral domain. Let $C=B[1/\bar f]\cong B[u]/(\bar f u-1)$, the localization and let $\mathfrak{m}$ be a maximal ideal of C. Note that C is the image of a surjective map $k[x_1,\dots,x_n,u]\to C$ and so $\mathfrak{m}$ comes from a maximal ideal $\mathfrak{m}^+$ of $k[x_1,\dots,x_n,u]$. The

weak nullstellensatz implies $k[x_1,\dots,x_n,u]/\mathfrak{m}^+ = k$ and we conclude that $C/\mathfrak{m} \cong k$. This means that there is a homomorphism of C onto k such that the image of $\bar{f}$ is not zero. Thus, the composition

$$k[x_1,\dots,x_n] \to B \to C \to C/\mathfrak{m} \to k$$

shows that f is not mapped to zero under this map, a map which assigns to the x_i some values λ_i. Now, under this map any element of $\mathfrak{p}$ is mapped to 0, which means that the point $(\lambda_1,\dots,\lambda_n) \in Z(\mathfrak{p}) \subseteq Z(\mathfrak{a})$, but $f(\lambda_1,\dots,\lambda_n) \neq 0$. Thus, $f \notin I(Z(\mathfrak{a}))$. □

Looking back at the proof we see that the main idea is to consider a point $(\lambda_1,\dots,\lambda_n)$ as a homomorphism of k-algebras, $k[x_1,\dots,x_n] \to k, x_i \mapsto \lambda_i$. Indeed, there is a bijection between points and such k-algebra homomorphisms; this is a reinterpretation of the Weak Nullstellensatz. We are therefore looking for such a homomorphism that will send the ideal $\mathfrak{p}$ to 0, guarantying that the point $(\lambda_1,\dots,\lambda_n)$ belongs to $Z(\mathfrak{p}) \subseteq Z(\mathfrak{a})$. Thus, by the first isomorphism theorem, such a homomorphism factors through B. To make sure that the homomorphism $B \to k$, that takes f to $f(\lambda_1,\dots,\lambda_n)$, does not send f to 0, we make f a unit, by passing to the localization C. At this point, the problem is to construct *any* homomorphism of k-algebras $C \to k$ and this is the step that was done by showing $C/\mathfrak{m} \cong k$.

7.4.2. *The classical affine space and* $\mathbf{Spec}(k[x_1,\dots,x_n])$. The message is that there is no real difference between the classical affine space and $\mathbf{Spec}(k[x_1,\dots,x_n])$ and so we see that the theory of spectra of rings *extends* classical algebraic geometry.

To every closed point $[\mathfrak{m}]$ of $\mathbf{Spec}(k[x_1,\dots,x_n])$ we associate a point of k^n: $\mathfrak{m}$ is a maximal ideal hence of the form $\langle x_1-\lambda_1,\dots,x_n-\lambda_n\rangle$ and we associate to it the point $(\lambda_1,\dots,\lambda_n)$. Every closed set of $\mathbf{Spec}(k[x_1,\dots,x_n])$ if of the form $V(\mathfrak{a})$ for $\mathfrak{a}$ a radical ideal, and it is assigned the algebraic set $Z(\mathfrak{a})$ in k^n. In particular, a point $[\mathfrak{p}]$ becomes in a sense an "ideal point" associated with the closed algebraic set $Z(\mathfrak{p})$. One can show that every irreducible algebraic set of k^n is of the form $V(\mathfrak{p})$ for a unique prime ideal $\mathfrak{p}$. One can say that the points of k^n, together with the architecture of algebraic sets is completely decribed by $\mathbf{Spec}(k[x_1,\dots,x_n])$. The points correspond to maximal ideals, algebraic sets to closed sets $V(\mathfrak{a})$, $\mathfrak{a}$ a radical ideal, and irreducible algebraic sets correspond to irreducible closed sets of $\mathbf{Spec}(k[x_1,\dots,x_n])$ that all have the form $V(\mathfrak{p}) = \overline{\{[\mathfrak{p}]\}}$ for a unique prime ideal. Thus, *"the algebra organizes the geometry."*

Part 3. REPRESENTATIONS OF FINITE GROUPS

In this manuscript we only consider finite groups G and finite dimensional complex vector spaces V. The theory of representations of infinite groups and infinite-dimensional representations is vast, and important, but is too advanced for this course. We should mention that even if one is interested in representations of Lie groups such as $\mathrm{GL}_n(\mathbb{C})$ or $U_n(\mathbb{C})$, which arise often in physics, the theory of representations of finite groups plays an important role.

Group representations are intimately related to understanding how groups act on sets. In our current setting, the set is a complex vector space and the group acts through very particular symmetries – invertible linear transformations. Thus, the topic of group representations can be viewed as a natural continuation of the combinatorial study of groups actions on sets.

Group representations are a subject with many applications to other branches of mathematics, and outside mathematics, for example for computer science, physics, chemistry, and electrical engineering. We will see some of those towards the end. It is also a topic that is a beautiful marriage of linear algebra and group theory, thus connecting two different areas of mathematics.

8. First definitions

A **linear representation** of a (finite) group G is a homomorphism

$$\rho\colon G \to \mathrm{GL}(V) := \{T\colon V \to V : T \text{ is an invertible linear transformation}\},$$

where V is a finite dimensional complex vector space. We will usually drop the adjective "linear". We note that $\mathrm{GL}(V)$ is a group under composition of linear maps. We will denote such a representation by (ρ, V), where the group G is understood from the context. When we feel confident enough, we may just denote it by either ρ, or V, depending which notation seems more useful at that point.

A very important notion is the notion of two representations being isomorphic. Given two representations (ρ_i, V_i) of G we define

$$\mathrm{Hom}_G(V_1, V_2) = \{T\colon V_1 \to V_2 \text{ linear} : T \circ \rho_1(g) = \rho_2(g) \circ T, \forall g \in G\}.$$

We note that there is no assumption that T is invertible, or even that $\dim(V_1) = \dim(V_2)$; in particular, we always have that the zero map is an element of $\mathrm{Hom}_G(V_1, V_2)$. Further, under addition of linear maps and multiplication by a scalar, $\mathrm{Hom}_G(V_1, V_2)$ is a complex vector space. We shall refer to elements of it as **homomorphisms of representations**, or G-**homomorphisms**.

Having made this definition, the notion of an **isomorphism** $(\rho_1, V_1) \cong (\rho_2, V_2)$ is clear: the isomorphisms are the linear maps $T \in \mathrm{Hom}_G(V_1, V_2)$ that are invertible. In that case, the inverse map T^{-1} always satisfies $T^{-1} \in \mathrm{Hom}_G(V_2, V_1)$.

Main Goal: Classify representations of G up to isomorphism

(We will make this more precise later on).

Given a representation (ρ, V), *choose* an isomorphism $T\colon V \to \mathbb{C}^n$ ($n = \dim(V)$) and let

$$\tau\colon G \to \mathrm{GL}(\mathbb{C}^n), \quad \tau(g) = T \circ \rho(g) \circ T^{-1}.$$

It is easily verified that

$$(\rho, V) \cong (\tau, \mathbb{C}^n),$$

where the isomorphism is the map T itself. Therefore, every isomorphism class of representations is represented by some $(\tau, \mathbb{C}^n)$.

How unique is τ? It is unique up to conjugation by elements of $\mathrm{GL}(\mathbb{C}^n)$: for any $T_1 \in \mathrm{GL}(\mathbb{C}^n)$ we have

$$\tau \cong \tau_1,$$

where

$$\tau_1(g) = T_1 \circ \tau(g) \circ T_1^{-1}.$$

(this reflects the fact that we had to *choose* an isomorphism $T\colon V \to \mathbb{C}^n$ and the freedom in this choice is precisely modifying T to $T_1 \circ T$).

It follows that we can make everything more concrete by using the natural identification

$$\mathrm{GL}(\mathbb{C}^n) = \mathrm{GL}_n(\mathbb{C}),$$

obtained by representing any linear transformation T by its matrix $[T]$ relative to the usual basis of $\mathbb{C}^n$. Thus, *we may think about a representation also as a homomorphism*

$$\tau\colon G \to \mathrm{GL}_n(\mathbb{C}).$$

The homomorphism rule is $\tau(xy) = \tau(x)\tau(y)$, where on the right we find matrix multiplication.

When do two such homomorphisms define isomorphic representations? For any invertible matrix $M \in \mathrm{GL}_n(\mathbb{C})$, we have

$$\tau \cong \rho, \qquad \rho(g) = M\tau(g)M^{-1}, \forall g \in G,$$

and conversely. This may be a confusing point, so let's repeat it: we are allowed to choose any matrix $M \in \mathrm{GL}_n(\mathbb{C})$ but, once we made the choice, the relation $\rho(g) = M\tau(g)M^{-1}$ should hold for all $g \in G$, *with the same M*.

Although we have finally arrived at a rather concrete model for representations, the general point of view $\rho\colon G \to \mathrm{GL}(V)$ is very useful as often the vector space V doesn't have a natural basis.

We now come to one of the key notions of this whole subject: the **character of a representation**. Given a representation

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

we define its **character** χ_ρ as follows:

$$\chi_\rho\colon G \to \mathbb{C}, \quad \chi_\rho(g) = \mathrm{Tr}(\rho(g)).$$

It is important to note that χ_ρ is simply a function; it associate to each element $g\mathcal{O}_M G$ the trace of the linear operator $\rho(g)$. Usually it will not have any multiplicative properties.

The notion of a character will turn out to be central for the whole theory and we will study many properties of characters. For now, we only give a few basic facts.

Lemma 8.0.1. *(1) χ_ρ only depends on the isomorphism class of ρ.*

(2) χ_ρ is constant on conjugacy classes in G.

(3) $\chi(1) = \dim(V)$.

Proof. To calculate the trace of an operator $\rho(g)$ one needs to choose a basis B for V and represent $\rho(g)$ by a matrix $[\rho(g)]_B$. If we choose another basis, say C, then the matrices of $\rho(g)$ in the two bases are related by

$$[\rho(g)]_C = M[\rho(g)]_B M^{-1},$$

where M is the change of basis matrix. Note that if we pass from ρ to an isomorphic representation, say (τ, W),

$$\tau(g) = T\rho(g)T^{-1}$$

then once more

$$[\tau(g)]_C = M[\rho(g)]_B M^{-1},$$

where now C is a basis of W and M is the matrix representing T relative to the two bases B, C. Thus, in both cases, we have to show that

$$\mathrm{Tr}(M[\rho(g)]_B M^{-1}) = \mathrm{Tr}([\rho(g)]_B).$$

This is well known (it follows from the formula $\mathrm{Tr}(MN) = \mathrm{Tr}(NM)$ that one proves by writing down the product of the matrices explicitly and calculating the trace).

The proof that χ_ρ is constant on conjugacy classes is very similar. Relative to some basis B we have

$$\mathrm{Tr}([\rho(hgh^{-1})]_B) = \mathrm{Tr}([\rho(h)\rho(g)\rho(h)^{-1}]_B) = \mathrm{Tr}([\rho(h)]_B[\rho(g)]_B[\rho(h)^{-1}]_B) = \mathrm{Tr}([\rho(g)]_B).$$

Finally, we have $\chi_\rho(1_G) = \mathrm{Tr}(\mathrm{Id}_V) = \mathrm{Tr}(I_{\dim(V)}) = \dim(V)$, where we denote by Id_V the identity operator on V and by I_d the $d \times d$ identity matrix. □

9. Examples

We now discuss some relatively simple examples. Despite appearances, perhaps, they will turn out to be very important and will make frequent appearances. Study them carefully!

9.1. 1-**dimensional representations.** A 1-dimensional representation of G could be thought of simply as a homomorphism

$$\rho\colon G \to \mathbb{C}^\times.$$

Indeed, $\mathrm{GL}_1(\mathbb{C}^\times) = \mathbb{C}^\times$. Note that in this case if $\rho \cong \tau$ then, since $\mathbb{C}^\times$ is commutative, we actually have $\rho = \tau$. Also, since the trace of a 1×1 matrix is (α) is just α it follows that

$$\chi_\rho = \rho.$$

For these reasons, 1-dimensional representations are also called 1-**dimensional characters**, or **multiplicative characters** .

Let

$$G^* = \mathrm{Hom}_{\mathbf{Gps}}(G, \mathbb{C}^\times).$$

We make two observations: First, G^* is a group under the rule

$$(\rho \cdot \tau)(g) = \rho(g) \cdot \tau(g).$$

Second, if we let $S^1 = \{z \in \mathbb{C}^\times : |z| = 1\}$ denote the unit circle in $\mathbb{C}$ then

$$G^* = \mathrm{Hom}(G, S^1).$$

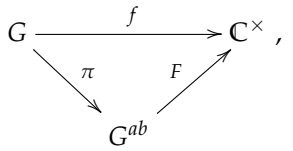
Indeed, if $g \in G$ is of order d, $\rho \in G^*$, then $\rho(g)^d = \rho(1_G) = 1$ which implies that $\rho(g)$ is necessarily a root of unity. The group G^* is called the **character group** of G.

Lemma 9.1.1. *There is a natural isomorphism*

$$G^* \cong (G^{ab})^*,$$

where, as usual, $G^{ab} = G/G'$ is the abelianization of G.

Proof. Any homomorphism $G \to A$, where A is an abelian group, factors uniquely through $G^{ab} = G/G'$, where G' is the commutator subgroup of G. In particular, given any homomorphism $f\colon G \to \mathbb{C}^\times$ there is a unique homomorphism $F\colon G^{ab} \to \mathbb{C}^\times$ such that the following diagram commutes (π being the natural map $G \to G/G'$):



and conversely. □

We will revisit this example later on, relying on the Exercises.

Example 9.1.2. The alternating groups A_n for $n \geq 5$ are simple and non-abelian hence $A_n^{ab} \cong \{1\}$ for $n \geq 5$. It follows that the alternating groups A_n for $n \geq 5$ have only one 1-dimensional representation, which is the trivial representation $\mathbb{1}$. For any group G the **trivial representation** $\mathbb{1}$ is the 1-dimensional representation

$$G \to \mathbb{C}^\times, \quad g \mapsto 1, \ \ \forall g \in G.$$

Its character, also denoted $\mathbb{1}$, is the constant function with value 1.

The symmetric groups S_n, for $n \geq 5$, have only two 1-dimensional characters, $\mathbb{1}$ and $\mathbf{sgn}$. Indeed, the only non-trivial normal subgroup of S_n, for $n \geq 5$, is A_n and, as $S_n/A_n \cong \{\pm 1\}$ is abelian, it must be that $S_n^{ab} \cong \{\pm 1\}$. The group $\{\pm 1\}$ has precisely two homomorphisms to $\mathbb{C}^\times$, the trivial one and the inclusion.

Example 9.1.3. The commutator subgroup of $D_4 := \langle x, y, | x^4 = y^2 = 1, yxy^{-1} = x^{-1} \rangle$ is $\{1, x^2\}$. Indeed, $[x, y] = x^2$ and so the commutator subgroup contains $\langle x^2 \rangle$. On the other hand, x^2 commutes with x and y and is therefore a central element and thus $\langle x^2 \rangle$ is a normal subgroup. As $D_4/\langle x^2 \rangle$ has order 2^2 it is abelian and it follows that $\langle x^2 \rangle \supseteq D_4'$. We conclude that $D_4' = \langle x^2 \rangle$. We think about the abelianization as

$$D_4^{ab} = \{1, \bar{x}, \bar{y}, \overline{xy}\}$$

with $\bar{x}\bar{y} = \bar{y}\bar{x}$ and the square of every element is 1; it is a group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$. As every element has order 2, every multiplicative character of D_4^{ab} takes values in $\{\pm 1\}$. It is not hard to show that there are 4 possibilities as described in the following table.

	1	$\bar{x}$	$\bar{y}$	$\overline{xy}$
$\rho_1 = \mathbb{1}$	1	1	1	1
ρ_2	1	-1	1	-1
ρ_3	1	1	-1	-1
ρ_4	1	-1	-1	1

9.2. **The regular representation ρ^{reg}.** Let G be a group. We defined a vector space with a basis $\{[g] : g \in G\}$, the group ring $\mathbb{C}[G]$ of G. A vector is a formal sum $\sum_{g\in G} a_g \cdot [g]$, where $a_g \in \mathbb{C}$.Recall that $\mathbb{C}[G]$ has a ring structure, where

$$(\sum_{g\in G} a_g \cdot [g]) + (\sum_{g\in G} b_g \cdot [g]) = \sum_{g\in G} (a_g + b_g) \cdot [g],$$

and

$$(\sum_{g\in G} a_g \cdot [g])(\sum_{g\in G} b_g \cdot [g]) = \sum_{g\in G} (\sum_{s\in G} a_{gs^{-1}} b_s) \cdot [g].$$

The group G acts on this vector space, giving the regular representation and denote it ρ^{reg}:

$$\rho^{reg}(g)(\sum_{s\in G} a_s[s]) = [g](\sum_{s\in G} a_s[s]) = \sum_{s\in G} a_s[gs].$$

The character χ^{reg} of ρ^{reg} is very simple:

$$\chi^{reg}(g) = \begin{cases} \sharp G, & g = 1_g; \\ 0, & else. \end{cases} \tag{7}$$

The proof is not hard: if $\{e_1, \dots, e_n\}$ is a basis for a vector space W, and $T\colon W \to W$ is a linear transformation, write

$$T(e_i) = \sum_{a=1}^{n} b_a e_a,\ b_a \in \mathbb{C}.$$

Then the contribution to $\mathrm{Tr}(T)$ from the vector e_i is b_i. Now, to calculate $\mathrm{Tr}(\rho^{reg}(g))$ we see that the contribution from the vector $[s]$ is the coefficient of $[s]$ in $\rho^{reg}(g)([s])$. As $\rho^{reg}(g)([s]) = [gs]$, this contribution is 0 from *every* s if $g \neq 1$, and is 1 from *every* s if $g = 1$.

9.3. **Direct sum.** Let $(\rho_1, V_1), (\rho_2, V_2)$ be two representations of the group G. We define the **direct sum** of the representations: the vector space is $V_1 \oplus V_2$ and the representation $\rho_1 \oplus \rho_2$ is as follows:

$$\rho_1 \oplus \rho_2\colon G \to \mathrm{GL}(V_1 \oplus V_2), \quad (\rho_1 \oplus \rho_2)(g)(v_1, v_2) := (\rho_1(g)(v_1), \rho_2(g)(v_2)).$$

If we represent ρ_i as homomorphisms,

$$\rho_i\colon G \to \mathrm{GL}_{n_i}(\mathbb{C}) \quad (n_i = \dim(V_i)),$$

then

$$\rho_1 \oplus \rho_2\colon G \to \mathrm{GL}_{n_1+n_2}(\mathbb{C}), \quad (\rho_1 \oplus \rho_2)(g) = \begin{pmatrix} \rho_1(g) & 0 \\ 0 & \rho_2(g) \end{pmatrix}.$$

This is a block diagonal matrix with the matrices $\rho_1(g), \rho_2(g)$ on the diagonal. It is then clear that

$$\chi_{\rho_1\oplus\rho_2}(g) = \chi_{\rho_1}(g) + \chi_{\rho_2}(g).$$

9.4. **Tensor product of representations.** Let (ρ_i, V_i) be representations of a group G. Then the **tensor product** representation has an underlying vector space

$$V_1 \otimes V_2,$$

where $g \in G$ acts by

$$(\rho_1(g) \otimes \rho_2(g))(v \otimes w) = \rho_1(v) \otimes \rho_2(w),$$

or more generally,

$$(\rho_1(g) \otimes \rho_2(g))(\sum_{i=1}^{n} v_i \otimes w_i) = \sum_{i=1}^{n} \rho_1(v_i) \otimes \rho_2(w_i).$$

The known properties of tensor products imply that this is indeed a linear presentation.

Let $A = (a_{ij})$ be an $m \times m$ complex matrix and $B = (b_{ij})$ an $n \times n$ complex matrix. The **Kronecker product** of A and B is the $mn \times mn$ matrix

$$A\mathbf{x}B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & \cdots \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{pmatrix}.$$

Let $v_1, \ldots, v_m$ be a basis for V_1 and $w_1, \ldots, w_n$ be a basis for V_2. Then,

$$v_1 \otimes w_1, \ldots, v_1 \otimes w_n,\ v_2 \otimes w_1, \ldots, v_2 \otimes w_n, \ldots,\ v_m \otimes w_1, \ldots v_m \otimes w_n,$$

(in that order!) is a basis for $V_1 \otimes V_2$. One check that if in the specified bases $[\rho_1(g)] = A$, $[\rho_2(g)] = B$, then

$$[\rho_1(g) \otimes \rho_2(g)] = A\mathbf{x}B.$$

We thus conclude:

$$\chi_{\rho_1 \otimes \rho_2} = \chi_{\rho_1} \cdot \chi_{\rho_2}.$$

9.5. **Induced and restricted representations.** Let G be a group and $H < G$ a subgroup. Given a representation (ρ, V) of G we get the **restricted representation**

$$\mathrm{Res}^G_H \rho := (\rho|_H, V).$$

That is, we simply restrict the homomorphism $\rho\colon G \to \mathrm{GL}(V)$ to H. This is a representation of H whose character is $\chi_\rho|_H$.

In the other direction, starting from a representation (ρ, V) of H, we have the **induced representation** $\mathrm{Ind}^G_H \rho$. The construction is more delicate.

$$\mathrm{Ind}^G_H \rho := \mathbb{C}[G] \otimes_{\mathbb{C}[H]} V.$$

The group ring $\mathbb{C}[G]$ acts by multiplication from the left making this into a representation of G. (We use the equivalence of the points of views "representations of G" $\leftrightarrow$ "$\mathbb{C}[G]$-modules" as in Example 1.3.3). We will later study this representation in more detail and calculate its character. It will turn out to be a very powerful method to construct representations of a group G from subgroups of it that are simpler. Note that even the trivial example, where $H = \{1\}$ and $(\rho, V) = (\mathbb{1}, \mathbb{C})$ is interesting – the induced representation of the trivial representation of the trivial subgroup is the regular representation of G.

10. Subrepresentations and irreducible representations

10.1. **Subrepresentions.** Let (ρ, V) be a representation of G. Let $U \subseteq V$ be a subspace such that

$$\rho(g)(u) \in U, \quad \forall g \in G, \forall u \in U.$$

That is, U is invariant under all the linear maps $\{\rho(g) : g \in G\}$. Then U is called a **subrepresentation** of V; we have

$$\rho|_U\colon G \to \mathrm{GL}(U), \quad \rho|_U(g) := \rho(g)|_U.$$

Example 10.1.1. $\{0\}$ and V are always sub-representations. We refer to them as **trivial subrepresentations.**

Example 10.1.2. The **standard representation** ρ^{std} of S_n.

Let $n \geq 2$. We consider S_n as contained in $\mathrm{GL}_n(\mathbb{C})$ in such a way that

$$\sigma(e_i) = e_{\sigma(i)}, \quad i = 1, 2, \ldots, n.$$

This is called the standard n-dimensional representation of S_n. For example, for $n = 3$,

$$(12) \leftrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (123) \leftrightarrow \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Let χ^{std} be the character of ρ^{std}. In our example of $n = 3$ we have $\chi^{std}(12) = 1, \chi^{std}(123) = 0$.

Proposition 10.1.3. *We have*

$$\chi^{std}(\sigma) = \sharp \text{ fixed points of } \sigma. \tag{8}$$

Proof. The contribution to $\mathrm{Tr}(\rho^{std}(\sigma))$ coming from the basis vector e_i is the coefficient of e_i in $\rho^{std}(\sigma)(e_i) = e_{\sigma(i)}$, which is 1 if $\sigma(i) = i$, and 0 if $\sigma(i) \neq i$. Summing over all i, we find the statement in the proposition. □

Consider now the subspaces

$$U_1 := \{(a, \ldots, a) : a \in \mathbb{C}\},$$

and

$$U_0 := \{(x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = 0, x_i \in \mathbb{C}\}.$$

The space U_1 is isomorphic to the trivial representation $\mathbb{1}$ of S_n, and U_0 is also a representation of S_n that we denote $\rho^{std,0}$. As $\dim(U_1) + \dim(U_0) = n$ and $U_1 \cap U_0 = \{0\}$, we find:

$$\rho^{std} = \mathbb{1} \oplus \rho^{std,0}. \tag{9}$$

10.2. **Irreducible representations and Maschke's Theorem.** A representation (ρ, V) of G is called **irreducible** if its only subrepresentations are $\{0\}$ and V, and $V \neq 0$.

Proposition 10.2.1. *The representations $\mathbb{1}$ and $\rho^{std,0}$ are irreducible representations of S_n. Thus, we have a decomposition of ρ^{std} as a sum of irreducible representations.*

Proof. Clearly $\mathbb{1}$ is irreducible for dimension reasons – there aren't any non-trivial subspaces; this is true for any group G and any 1-dimensional representation of it.

The proof for U_0 is slightly involved; we will give another proof later, much more elegant, as an application of character theory.

We assume that $n > 2$. The case $n = 2$ is easy as U_0 is 1-dimensional.

Let $U' \subseteq U_0$ be a non-zero sub-representation. Let $x = (x_1, \ldots, x_n)$ be a non-zero vector in U'. If x has precisely two zero elements, by multiplying x by a scalar we may assume that $x = (0, \ldots, 0, 1, 0 \ldots, 0, -1, 0, \ldots, 0)$. Then, by acting by S_n we see that every vector of the form $e_i - e_j$ (where e_i are the standard basis) is also in U'. But these vectors span U_0 and it follows that $U' = U_0$.

Thus, it remains to prove that U' always contains such a vector. Let $x \in U'$ be a non-zero vector. If x has more than 2 non-zero coordinates, we show that there is vector $y \in U'$ that is not zero and has fewer non-zero coordinates. This suffices to reduce to the case considered above.

Assume therefore that x has at least 3 non-zero coordinates. First, by rescaling we may assume that one of these coordinates is 1. Then, as $\sum x_i = 0$, there exists a non-zero coordinate that is not equal to 1. By applying a permutation to x we may assume that

$$x = (1, x_2, x_3, \ldots, x_n),$$

where $x_2 \neq 1$ and is non-zero and also $x_3 \neq 0$. In this case, also the vector

$$x' = \frac{1}{x_2}(x_2, 1, x_3, \ldots, x_n),$$

belongs to U'. Therefore, also

$$y = x - x' = (0, x_2 - \frac{1}{x_2}, x_3(1 - \frac{1}{x_2}), \ldots, x_n(1 - \frac{1}{x_2})),$$

belongs to U' and this vector has fewer non-zero coordinates, yet is not zero (consider its third coordinate). □

Theorem 10.2.2 (Maschke)**.** *Every non-zero representation (ρ, V) decomposes as a direct sum of irreducible representations.*

Remark 10.2.3. We will later prove that such a direct sum decomposition is unique, up to isomorphism and re-ordering of the summands. We can now make our goal in this manuscript more precise:

> **Main Goal:** Classify the irreducible representations of a group G. Find effective methods to determine the decomposition of a representation into irreducible representations.

Proof. (Maschke's Theorem) We begin with a lemma that shows that we can always define an inner product on V relative to which $\rho(g)$ is a unitary matrix for any $g \in G$.

Lemma 10.2.4. *There is an inner product*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C},$$

such that

$$\langle gv, gu \rangle = \langle v, u \rangle, \quad \forall g \in G, \forall u, v \in V.$$

(To simplify notation we write gv for $\rho(g)(v)$.)

Proof. (Lemma) Let $(\cdot, \cdot)$ be *any* inner product on V. Define,

$$\langle v, u \rangle = \frac{1}{\sharp G} \sum_{g \in G} (gv, gu).$$

The verification that this is an inner product is straightforward and we omit it. To check that ρ is a unitary representation relative to this inner product we calculate:

$$\begin{aligned} \langle gv, gu \rangle &= \frac{1}{\sharp G} \sum_{h \in G} (hgv, hgu) \\ &= \frac{1}{\sharp G} \sum_{h \in G} (hv, hu) \\ &= \langle v, u \rangle, \end{aligned}$$

where we used that when h runs over G so does hg. □

We now get to the proof of the theorem. We prove it by induction on $\dim(V)$.

If $\dim(V) = 1$ then V is irreducible and there is nothing to prove. In general, if V is irreducible there is nothing to prove. Otherwise, V has a subrepresentation $0 \neq U \neq V$. Let $\langle v, u\rangle$ be a G-invariant inner product on V, as in the Lemma. Then

$$V = U \oplus U^{\perp}.$$

We show that

$$U^{\perp} := \{v \in V : \langle v, u\rangle = 0, \forall u \in U\}$$

is a subrepresentation. Let $g \in G$ and $v \in U^{\perp}$. For any $u \in U$ we have

$$\langle gv, u\rangle = \langle v, g^{-1}u\rangle = 0,$$

because $g^{-1}u \in U$ as U is a subrepresentation. It follows that $gv \in U^{\perp}$.

By induction,

$$U = W_1 \oplus \cdots \oplus W_a, \quad U^{\perp} = W_{a+1} \oplus \cdots \oplus W_b,$$

for some irreducible representations W_i of G. Then,

$$V = U \oplus U^{\perp} = W_1 \oplus \cdots \oplus W_b$$

is a sum of irreducible representations too. □

10.3. **The projection on V^G.** Let (ρ, V) be a representation of G. Let

$$V^G = \{v \in V : \rho(g)(v) = v, \forall g \in G\}.$$

Then V^G is a subrepresentation on which G acts trivially. It's the space of **invariant vectors**.

Lemma 10.3.1. *Let*

$$\pi(v) = \frac{1}{\sharp G} \sum_{g \in G} \rho(g)(v). \tag{10}$$

Then $\pi \in \mathrm{Hom}_G(V, V^G)$ and is a projection on the subspace V^G.

Proof. As π is a sum of linear maps it is certainly a linear map from V to V. We first show that $\mathrm{Im}(\pi) \subseteq V^G$. We need to show that all $h \in G, v \in V$ we have $\rho(h)(\pi(v)) = \pi(v)$. Indeed, $\rho(h)(\pi(v)) = \frac{1}{\sharp G}\sum_g (\rho(h) \circ \rho(g))(v) = \frac{1}{\sharp G}\sum_g \rho(hg)(v) = \frac{1}{\sharp G}\sum_g \rho(g)(v) = \pi(v)$.

To show π is a projection, we need to verify that π is the identity on V^G. But, for $v \in V^G$ we have $\pi(v) = \frac{1}{\sharp G}\sum_g \rho(g)(v) = \frac{1}{\sharp G}\sum_g v = v$.

Finally, we check that π is a homomorphism of representations. As G acts trivially on V^G this boils down to verifying that $\pi(\rho(h)v) = \pi(v)$. We calculate: $\pi(\rho(h)(v)) = \frac{1}{\sharp G}\sum_g \rho(g)(\rho(h)v) = \frac{1}{\sharp G}\sum_g \rho(gh)(v) = \frac{1}{\sharp G}\sum_g \rho(g)(v) = \pi(v)$. □

The following corollary will be used several times in the sequel:

Corollary 10.3.2 (Projection Formula)**.** *We have*

$$\dim(V^G) = \frac{1}{\sharp G}\sum_g \chi_\rho(g). \tag{11}$$

In words, the dimension of the subspace of invariant vectors is the average value of the character χ_ρ on the group G.

Proof. We have a decomposition,

$$V = V^G \oplus \mathrm{Ker}(\pi).$$

In this decomposition we can write

$$\pi = \mathrm{Id}_{V^G} \oplus 0.$$

Thus, $\mathrm{Tr}(\pi) = \dim(V^G)$. On the other hand,

$$\mathrm{Tr}(\pi) = \frac{1}{\sharp G}\sum_g \mathrm{Tr}(\rho(g)) = \frac{1}{\sharp G}\sum_g \chi_\rho(g).$$

□

Example 10.3.3. The action of S_3 on itself by multiplication from the left, as in Cayley's Theorem, provides us with an embedding $S_3 \hookrightarrow S_6$. Composing with the standard representation of S_6 we get a 6-dimensional representation ρ of S_3. Does this representation have fixed vectors? what is the dimension of the space of fixed vectors??

If we enumerate the elements of S_3 as $\{1,(12),(13),(132),(23),(123)\} = \{\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_5,\sigma_6\}$, then

$$\rho((12)) = \begin{pmatrix} 0&1&0&0&0&0\\ 1&0&0&0&0&0\\ 0&0&0&1&0&0\\ 0&0&1&0&0&0\\ 0&0&0&0&0&1\\ 0&0&0&0&1&0 \end{pmatrix}, \qquad \rho((123)) = \begin{pmatrix} 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&1&0&0&0&0\\ 0&0&0&0&0&1\\ 0&0&1&0&0&0\\ 1&0&0&0&0&0 \end{pmatrix}.$$

(For example, $(12) \in S_3$ takes to σ_1 to σ_2, σ_2 to σ_1, σ_3 to σ_4, etc. and so corresponds to the permutation $(12)(34)(56) \in S_6$.)

If χ denotes the character of ρ, then by calculating χ on $1, (12)$ and (123) we would know its value on any $\sigma \in S_3$, because a character has a fixed value on each conjugacy class. We find

$$\chi(1) = 6, \quad \chi((ij)) = 0, \quad \chi((ijk)) = 0.$$

It follows from Corollary 10.3.2 that the dimension of the space of invariant vectors is 1.

Finally, note that we could have found the values of χ without writing down the matrices. Just by observation we could say that any non-identity element of S_3 is mapped to a permutation in S_6 that has no fixed points.[20] As for a permutation $\sigma \in S_6$, the value of $\chi^{std}(\sigma) = \sharp(\text{fixed points of } \sigma)$, it follows that $\chi(\tau) = 0$ for any $1 \neq \tau \in S_3$.

Example 10.3.4. Let $(\rho, V) = (\rho^{std}, \mathbb{C}^n)$ be the standard representation. Then

$$\pi = \frac{1}{n!}\sum_{\sigma \in S_n} \rho^{std}(\sigma).$$

One checks that $V^G = U_1$ and $\mathrm{Ker}(\pi) = U_0$ (for the latter, it is easier to show $\mathrm{Ker}(\pi) \supseteq U_0$ and deduce equality by comparing dimensions). We find again the decomposition (9):

$$\rho^{std} = \mathbb{1} \oplus \rho^{std,0}.$$

Moreover, we find that

$$1 = \dim(U_1) = \dim(V^G) = \frac{1}{n!}\sum_{\sigma \in S_n} \sharp \text{ fixed points of } \sigma,$$

a formula one can also derive from the Cauchy-Frobenius formula. This is nice! It says that the average number (or expected number) of fixed points for a permutation chosen randomly from S_n is 1.

[20] That would be true for any group! If G has n elements, acting on itself by left multiplication which induces the embedding $G \hookrightarrow S_N$ in the proof of Cayley's theorem, then any $1 \neq g \in G$ acts without fixed points.

11. Schur's lemma and orthogonality of characters

11.1. **The dual representation and the two Homs.** Let (ρ, V) be a representation of G. For any linear operator $\rho(g)\colon V \to V$ we have the dual operator $\rho(g)^t\colon V^* \to V^*$, where $V^* = \mathrm{Hom}_{\mathbb{C}}(V, \mathbb{C})$ is the dual vector space to V. Recall that $\rho(g)^t$ is defined by[21]

$$\rho(g)^t(\varphi) = \varphi \circ \rho(g), \quad \varphi \in V^*.$$

Further, if $\{v_1, \dots, v_n\}$ are a basis for V and $\{v_1^*, \dots, v_n^*\}$ is the dual basis for V (the basis that satisfies $v_i^*(v_j) = \delta_{ij}$) then in terms of matrices we have

$$[\rho(g)^t]_{\{v_i^*\}} = ([\rho(g)]_{\{v_i\}})^t.$$

Define the **dual representation** ρ^*

$$\rho^*\colon G \to \mathrm{GL}(V^*), \quad \rho^*(g) = (\rho(g^{-1}))^t.$$

Proposition 11.1.1. *ρ^* is a representation of G and its character satisfies $\chi_{\rho^*} = \bar{\chi}_\rho$. That is,*

$$\chi_{\rho^*}(g) = \overline{\chi}_\rho(g) := \overline{\chi_\rho(g)}, \ \ \forall g \in G.$$

Proof. The proof is easy, but reveals two properties that are very important, and general, and so we record them here in a lemma.

Lemma 11.1.2. *Let (ρ, V) be a representation of G. Then:*

1. *Every $\rho(g)$ is diagonalizable.*
2. *Every eigenvalue of $\rho(g)$ is a root of unity of order dividing d, where d is the order of g in G.*

Proof. Let d be the order of g. As ρ is a homomorphism $\rho(g)^d = \rho(g^d) = \rho(1_G) = \mathrm{Id}_V$. It follows that $\rho(g)$ solves the polynomial $x^d - 1$, which is a separable polynomial (i.e., it has distinct roots over $\mathbb{C}$). Therefore, also the minimal polynomial of $\rho(g)$ is a separable polynomial and, consequently, $\rho(g)$ is diagonalizable. Also the second statement is now clear. □

Let's write

$$\rho(g) \sim \mathrm{diag}(\alpha_1, \dots, \alpha_n),$$

where $n = \dim(V)$ and α_i are d-th roots of unity. Note that in general the basis in which $\rho(g)$ is diagonal depends on g; we cannot, in general, diagonalize all the operators $\rho(g)$ simultaneously. However, $\rho(g^{-1}) = \rho(g)^{-1}$ is given in the same basis by

$$\mathrm{diag}(\alpha_1^{-1}, \dots, \alpha_n^{-1}) = \mathrm{diag}(\overline{\alpha_1}, \dots, \overline{\alpha_n}),$$

because the α_i are roots of unity. Thus,

$$\chi_{\rho^*}(g) = \chi_\rho(g^{-1}) = \sum_i \overline{\alpha_i} = \overline{\chi_\rho(g)}. \tag{12}$$

To finish the proof of the Proposition it only remains to check that ρ^* is a representation. We have:

$$\rho^*(gh) = (\rho(gh)^{-1})^t = (\rho(h^{-1})\rho(g^{-1}))^t = (\rho(g^{-1}))^t \cdot (\rho(h^{-1}))^t = \rho^*(g) \cdot \rho^*(h).$$

□

[21] We previously denoted the transformation dual to a linear map $T\colon V \to W$ by $T^*\colon W^* \to V^*$, where $T^*\varphi = \varphi \circ T$; that is, for any $v \in V$, $(T^*\varphi)(v) := \varphi(Tv)$, we now denote it T^t to hopefully avoid confusion in the definition of the dual representation to ρ.

We now discuss "the two Homs" and engage in a very technical calculation. However, the results will be absolutely essential to proving one of the most important theorems concerning representations: orthogonality of characters.

Let $(\rho, V), (\tau, W)$ be two representations of the group G. We have already defined (all maps appearing below are understood to be linear) the following vector space:

$$\mathrm{Hom}_G(V,W) = \{T\colon V \to W : T \circ \rho(g) = \tau(g) \circ T, \forall g \in G\}.$$

We also have the more naive

$$\mathrm{Hom}(V,W) = \{T\colon V \to W\}.$$

Proposition 11.1.3. $\mathrm{Hom}(V,W)$ *is a linear representation* σ *of* G*, where*

$$\sigma(g)(T) = \tau(g) \circ T \circ \rho(g)^{-1}, \quad T \in \mathrm{Hom}(V,W).$$

Remark 11.1.4. Note the following:

(1) $\dim(\mathrm{Hom}(V,W)) = \dim(V) \cdot \dim(W)$. This can be seen by choosing bases for the two vector spaces and representing the linear maps as matrices.

(2) We have the following relationship between the two Homs:

$$\mathrm{Hom}_G(V,W) = \mathrm{Hom}(V,W)^G.$$

(3) Consider the special case where $(\tau, W) = (\mathbb{1}, \mathbb{C})$. In this case

$$\mathrm{Hom}(V,W) = V^*,$$

and the new representation σ we have now defined is:

$$\sigma(g)(\phi) = \tau(g) \circ \phi \circ \rho(g^{-1}) = \phi \circ \rho(g^{-1}) = \rho(g^{-1})^t(\phi) = \rho^*(\phi).$$

Namely, we just get the dual representation again.

Proof. (Proposition) Let us generalize the last observation. Recall the isomorphism

$$V^* \otimes W \cong \mathrm{Hom}(V,W).$$

If $\{v_i\}$ are a basis for V, $\{v_i^*\}$ the dual basis for V^* and $\{w_i\}$ are a basis for W, then under this isomorphism, for any complex scalars a_{ij},

$$\sum_{ij} a_{ij} \cdot v_i^* \otimes w_j \mapsto T \in \mathrm{Hom}(V,W), \quad T(v) = \sum_{ij} a_{ij} v_i^*(v) \cdot w_j.$$

Now recall the construction of the tensor product representation. If $g \in G$ acts on v_i^* by $\rho(g^{-1})^t v_i^*$ (which means that $(\rho(g^{-1})^t v_i^*)(v) = v_i^*(\rho(g^{-1})v)$) and on w_i by $\tau(g)(w_i)$ then in the tensor product representation we have

$$(\rho(g^{-1})^t \otimes \tau(g))(\sum_{ij} a_{ij} v_i^* \otimes w_j) = \sum_{ij} a_{ij} (v_i^* \circ \rho(g^{-1})) \otimes \tau(g)(w_j),$$

which is sent to the linear transformation

$$v \mapsto \sum_{ij} a_{ij} v_i^*(\rho(g^{-1})(v)) \tau(g)(w_j) = \tau(g) \left(\sum_{ij} a_{ij} \cdot v_i^*(\rho(g^{-1})(v)) w_j \right).$$

But this is exactly the linear transformation $\tau(g) \circ T \circ \rho(g^{-1})$. This proves that $\mathrm{Hom}(V,W)$ is indeed a representation of G and in fact also proves the following lemma. $\square$

Lemma 11.1.5. *As representations of the group G,*

$$\mathrm{Hom}(V,W) \cong V^* \otimes W$$

We conclude the following theorem:

Theorem 11.1.6. *The character χ_σ of the representation $(\sigma, \mathrm{Hom}(V,W))$ is given by the formula*

$$\chi_\sigma = \chi_\tau \cdot \bar{\chi}_\rho$$

Proof. It is enough to calculate the character of $V^* \otimes W$, which is a formal computation at this point:

$$\chi_{\rho^* \otimes \tau} = \chi_{\rho^*} \cdot \chi_\tau = \chi_\tau \cdot \bar{\chi}_\rho.$$

□

11.2. **Schur's Lemma.** Before proving Schur's lemma, we establish some general properties of homomorphisms of representations.

Lemma 11.2.1. *For any two representations $(\rho, V), (\tau, W)$ of G and any $T \in \mathrm{Hom}_G(V,W)$ we have that $\mathrm{Ker}(T)$ is a subrepresentation of V, and $\mathrm{Im}(T)$ is a subrepresentation of W.*

Proof. Let $v \in \mathrm{Ker}(T)$ and $g \in G$. We have

$$T(\rho(g)(v)) = \tau(g)(T(v)) = \tau(g)(0) = 0.$$

It follows that $\mathrm{Ker}(T)$ is a subrepresentation of V.

Let $w \in \mathrm{Im}(T)$ and choose $v \in V$ such that $T(v) = w$. Then:

$$\tau(g)(w) = \tau(g)(T(v)) = T(\rho(g)(v)) \in \mathrm{Im}(T).$$

It follows that $\mathrm{Im}(T)$ is a subrepresentation of W. □

Lemma 11.2.2 (Schur)**.** *Let $(\rho, V), (\tau, W)$ be two irreducible representations of G. Then*

$$\mathrm{Hom}_G(V,W) \cong \begin{cases} \mathbb{C}, & (\rho,V) \cong (\tau,W); \\ 0, & \textit{else.} \end{cases} \tag{13}$$

Proof. Let $T \in \mathrm{Hom}_G(V,W)$ and suppose $T \neq 0$. Then $\mathrm{Ker}(T) \neq V$. However, $\mathrm{Ker}(T)$ is a subrepresentation of V and V is irreducible. It follows that $\mathrm{Ker}(T) = 0$ and so that T is injective. Since V is not zero (by definition) $\mathrm{Im}(T) \neq 0$, and since W is irreducible and $\mathrm{Im}(T)$ is a subrepresentation, $\mathrm{Im}(T) = W$. Thus, T is surjective. It follows that T is an isomorphism. Therefore, if $\mathrm{Hom}_G(V,W) \neq 0$ (and V, W are irreducible) we have $(\rho, V) \cong (\tau, W)$.

It remains to show that if $(\rho, V) \cong (\tau, W)$ then $\mathrm{Hom}_G(V,W)$ is a 1-dimensional vector space. Choose, any non-zero $T \in \mathrm{Hom}_G(V,W)$. We saw that T is then an isomorphism. We get an isomorphism

$$\mathrm{Hom}_G(V,W) \cong \mathrm{End}_G(V), \quad S \mapsto T^{-1} \circ S,$$

and thus it is enough to prove that

$$\mathrm{End}_G(V) \cong \mathbb{C}.$$

Let then $R \in \mathrm{End}_G(V)$ and let λ be an eigenvalue of R. As $\lambda \cdot \mathrm{Id} \in \mathrm{End}_G(V)$, it follows that $R - \lambda \cdot \mathrm{Id} \in \mathrm{End}_G(V)$ and it follows that $\mathrm{Ker}(R - \lambda \cdot \mathrm{Id})$ is a subrepresentation of V. Since every eigenvalue has at least one non-zero eigenvector, we have that $\mathrm{Ker}(R - \lambda \cdot \mathrm{Id}) \neq 0$ and, as V is irreducible, we must have

$$\mathrm{Ker}(R - \lambda \cdot \mathrm{Id}) = V.$$

This means that $R = \lambda \cdot \mathrm{Id}$. Conversely, $\lambda \cdot \mathrm{Id}$ always belongs to $\mathrm{End}_G(V)$ (for any representation (ρ, V) whatsoever). This provides the isomorphism $\mathrm{End}_G(V) \cong \mathbb{C}$. □

Remark 11.2.3. Note that the final isomorphism $\mathrm{End}_G(V) \cong \mathbb{C}$ can be given by

$$R \mapsto \frac{1}{\dim(V)} \cdot \mathrm{Tr}(R). \tag{14}$$

11.3. **The space of class functions.** Let G be a finite group and denote by $h(G)$ the **class number** of G. By definition, $h(G)$ is the number of conjugacy classes in G.

Example 11.3.1.
- If G is abelian, $h(G) = \sharp G$.
- If $G = S_n$, $h(G) = p(n)$ (the partition function of n).

A function $f\colon G \to \mathbb{C}$ is called a **class function** if

$$f(hgh^{-1}) = f(g), \quad \forall g, h \in G.$$

Namely, if f is constant on each conjugacy class. We let $\mathrm{Class}(G)$ denote the space of class functions. It is a complex vector space of dimension $h(G)$. If $\phi \in \mathrm{Class}(G)$, define a function $\bar{\phi} \in \mathrm{Class}(G)$ by

$$\bar{\phi}(g) := \overline{\phi(g)}$$

(where on the right we are simply taking the complex conjugate of the complex number $\phi(g)$).

We make $\mathrm{Class}(G)$ into a hermitian space by defining an inner product on it:

$$\langle \phi, \psi \rangle := \frac{1}{\sharp G} \sum_{g \in G} \phi(g) \cdot \bar{\psi}(g).$$

It is easy to verify that this is an inner product; we leave that as an exercise. We also define $\|\phi\|$ to be the non-negative real number satisfying $\|\phi\|^2 := \langle \phi, \phi \rangle$. Our main motivation is the following key example.

Example 11.3.2. For any representation (ρ, V) of G, its character $\chi_\rho \in \mathrm{Class}(G)$.

Example 11.3.3. Let $1 \leq r \leq n$ be integers. Define $\phi_r\colon S_n \to \mathbb{C}$ by $\phi(\sigma)$ equal to the number of cycles of length r appearing in the decomposition of σ as a product of disjoint cycles. The function ϕ_r is a class function.

While $\phi_1(\sigma)$ is the number of fixed points of σ and so $\phi_1 = \chi^{std}$, for $r > 1$ the function ϕ_r does not arise as a character of a representation. Indeed, $\phi_r(1) = 0$ for $r > 1$ so such a representation would have to be 0-dimensional, but for $r \leq n$ the function ϕ_r is not zero: $\phi_r((12 \cdots r)) = 1$.

11.4. **Orthogonality of characters.** We now come to the theorem making characters into a very powerful tool in the study of representations.

Theorem 11.4.1 (Orthogonality of characters)**.** *Let $(\rho, V), (\tau, W)$ be two irreducible representations of G. Then:*

(1) If $\rho \ncong \tau$ then $\langle \chi_\rho, \chi_\tau \rangle = 0$.
(2) $\|\chi_\rho\| = 1$.

Otherwise said, the characters of the irreducible representations of a group G form an orthonormal set in the space of class functions Class(G).

Remark 11.4.2. We will prove in Theorem 15.1.1 below that, in fact, the characters of irreducible representations form an orthonormal *basis* for Class(G).

Proof. Let us write $U = \mathrm{Hom}(V,W)$. We have seen that (σ, U) is a representation of G, where

$$\sigma\colon G \to \mathrm{GL}(U), \quad \sigma(g)(T) = \tau(g) \circ T \circ \rho(g^{-1}),$$

and, by Theorem 11.1.6,

$$\chi_\sigma = \chi_\tau \cdot \bar{\chi}_\rho.$$

By Schur's Lemma,

$$\dim(U^G) = \dim(\mathrm{Hom}_G(V,W)) = \begin{cases} 1, & \rho \cong \tau; \\ 0, & \rho \ncong \tau. \end{cases}$$

On the other hand, by the Projection Formula (Corollary 10.3.2), we have

$$\dim(U^G) = \frac{1}{\sharp G} \sum_{g \in G} \chi_\sigma(g) = \frac{1}{\sharp G} \sum_{g \in G} \chi_\tau(g) \cdot \bar{\chi}_\rho(g) = \langle \chi_\rho, \chi_\tau \rangle.$$

The theorem follows. □

Corollary 11.4.3. *Let h be the number of irreducible characters of G, up to isomorphism. We have*

$$h \leq h(G).$$

In words, the number of irreducible representations of G is at most its class number. (We will see later that $h = h(G)$.)

The following notation will be used repeatedly. Let

$$\rho_1, \dots, \rho_h,$$

be representatives to the isomorphism classes of irreducible representations of G. More precisely, we should say, let $\{(\rho_i, V_i) : i = 1, \dots, h\}$ be representatives to the isomorphism classes of irreducible representations of G, but this is heavier notation that we will usually avoid. In the same vain, given a representation (ρ, V) instead of saying that

$$(\rho, V) \cong (\rho_1, V_1)^{\oplus a_1} \oplus \cdots \oplus (\rho_h, V_h)^{\oplus a_h},$$

we will simply write

$$\rho \cong \rho_1^{a_1} \oplus \cdots \oplus \rho_h^{a_h}.$$

(Here the a_i are non-negative integers and the notation $(\rho_1, V_1)^{\oplus a_1}$ means the direct sum of (ρ_1, V_1) with itself a_1 times, which is declared to be the zero vector space 0 if $a_1 = 0$.) We will also use the notation

$$d_i = \dim(\rho_i), \quad \chi_i = \chi_{\rho_i}.$$

Finally, whenever we view ρ_i as homomorphisms

$$\rho_i\colon G \to \mathrm{GL}_{d_i}(\mathbb{C}),$$

we will assume, when convenient, that $\{\rho_i(g) : g \in G\}$ are *unitary* matrices, which can always be arranged, as we have seen while proving Maschke's theorem.

11.5. **Unique decomposition.** We now prove that the decomposition provided by Maschke's theorem is unique.

Theorem 11.5.1. *Let ρ be a representation of G. Then there are unique non-negative integers m_i such that*

$$\rho \cong \rho_1^{m_1} \oplus \cdots \oplus \rho_h^{m_h}.$$

Proof. By Maschke's theorem, such m_i always exist. Then, by using the formula for the character of a direct sum (§9.3), we have

$$\chi_\rho = \sum_{i=1}^{h} m_i \cdot \chi_i.$$

On the other hand, we can use this formula to deduce by orthogonality of characters that

$$\langle \chi_\rho, \chi_j \rangle = \langle \sum_{i=1}^{h} m_i \cdot \chi_i, \chi_j \rangle = m_j.$$

That shows that the multiplicities m_i are determined uniquely by ρ. □

We will refer to the m_i as the **multiplicity** of the irreducible representation ρ_i in ρ.

Corollary 11.5.2. *We have an isomorphism $(\rho, V) \cong (\tau, W)$ if and only if $\chi_\rho = \chi_\tau$. In words, the isomorphism class of a representation is completely determined by its character.*

Proof. One of the first properties of characters we proved was that the character depends only on the isomorphism class. So, the "only if" is clear. Suppose now that $\chi_\rho = \chi_\tau$, then for every χ_j we have $\langle \chi_\rho, \chi_j \rangle = \langle \chi_\tau, \chi_j \rangle =: m_j$. We have seen that then both representations are isomorphic to $\rho_1^{m_1} \oplus \cdots \oplus \rho_h^{m_h}$, hence to each other. □

12. Further theorems and examples

Before proving some additional "big theorems", we study some examples and prove some easier results that will give us a better sense of the whole subject.

12.1. **Decomposition of the regular representation.** Recall from § 9.2 the regular representation ρ^{reg} of a group G. It is the representation on the vector space $\mathbb{C}[G]$ that has basis $\{[g] : g \in G\}$, and

$$\rho^{reg}(h)([g]) = [hg], \quad \forall g, h \in G.$$

We have calculated there that

$$\chi^{reg}(g) = \begin{cases} \sharp G, & g = 1_G; \\ 0, & \text{else.} \end{cases}$$

Let us now find the decomposition of the regular representation into irreducible representations. As we have seen, the multiplicity m_i of χ_i is given by

$$m_i = \langle \chi^{reg}, \chi_i \rangle.$$

This is easy to calculate:

$$\langle \chi^{reg}, \chi_i \rangle = \frac{1}{\sharp G} \sum_g \chi^{reg}(g) \cdot \bar{\chi}_i(g) = \frac{1}{\sharp G} \cdot \chi^{reg}(1_g) \cdot \bar{\chi}_i(1_g) = d_i,$$

where $d_i = \dim(V_i)$, as per our conventions. We conclude the following proposition.

Proposition 12.1.1. *We have*

$$\rho^{reg} \cong \oplus_{i=1}^{h} \rho_i^{d_i}, \quad \chi^{reg} = \sum_{i=1}^{h} d_i \chi_i. \tag{15}$$

Namely, every irreducible representation appears in the regular representation with multiplicity equal to its dimension.

By calculating the dimensions of both sides in the isomorphism (15), we conclude:

Corollary 12.1.2. *We have*

$$\sharp G = \sum_{i=1}^{h} d_i^2. \tag{16}$$

12.2. **Criterion for being irreducible.** An easy consequence of orthogonality of characters is the following useful result.

Corollary 12.2.1. *A representation (ρ, V) is irreducible if and only if*

$$\|\chi_\rho\| = 1.$$

Proof. Let us write

$$\chi_\rho = \sum_i m_i \cdot \chi_i,$$

for non-negative integers m_i. By orthogonality of characters (Pythagoras), we have

$$\|\chi_\rho\|^2 = \sum_i m_i^2.$$

Thus, $\|\chi_\rho\| = 1$ if and only if there exists a unique i_0 such that $m_{i_0} = 1$ and all the rest of 0. But this is exactly the case where ρ is irreducible. □

Remark 12.2.2. A very similar argument gives that $\|\chi_\rho\|^2 = 2$ if and only if ρ is a sum of two distinct irreducible representations, and that $\|\chi_\rho\|^2 = 3$ if and only if ρ is a sum of three distinct irreducible representations. However, when $\|\chi_\rho\|^2 = 4$ the pattern breaks down, and ρ could be either the sum of four distinct irreducible representations, or isomorphic to two copies of a single irreducible representation.

12.3. **Another look at the standard representation of S_n.** We take another look here at the standard representation of S_n, $n \geq 2$, introduced in Example 10.1.2. Recall that this is an n-dimensional representation ρ^{std} of S_n whose character χ^{std} satisfies

$$\chi^{std}(\sigma) = I(\sigma) = \sharp \text{ fixed points of } \sigma.$$

We repeat here a calculation we have done before: It is clear that the space of invariant vectors is $(\mathbb{C}^n)^{S_n} = U_1$ in the notation of that example and, in particular, $\dim((\mathbb{C}^n)^{S_n}) = 1$. The projection formula gives another way to calculate this dimension (see Example 10.3.4) and we get

$$\frac{1}{n!} \sum_{\sigma \in S_n} \chi^{std}(\sigma) = \frac{1}{n!} \sum_{\sigma \in S_n} I(\sigma) = 1.$$

This has a pleasant interpretation: the expected number of fixed points for a randomly chosen permutation is 1.

Let us use the notation $T = \{1, 2, \dots, n\}$. Then, from the very definition of the inner product, we can say that

$$\|\chi^{std}\|^2 = \frac{1}{n!} \sum_{\sigma \in S_n} (\sharp \text{ fixed points of } \sigma \text{ on } T)^2.$$

Lemma 12.3.1. $\|\chi^{std}\|^2 = 2$.

Proof. Consider the action of S_n on $T \times T$ given by

$$\sigma(i, j) = (\sigma(i), \sigma(j)).$$

It is clear that S_n has two orbits on $T \times T$. Namely, $\{(i, i) : i \in T\}$ and $\{(i, j) : i \neq j \in T\}$. On the other hand, σ fixes (i, j) if and only if $\sigma(i) = i$ and $\sigma(j) = j$. Thus,

$$\sharp \text{ fixed points of } \sigma \text{ on } T \times T = (\sharp \text{ fixed points of } \sigma \text{ on } T)^2.$$

We apply the CFF to the action of S_n on $T \times T$ to conclude that

$$2 = \frac{1}{n!} \sum_{\sigma} \sharp \text{ fixed points of } \sigma \text{ on } T \times T = \frac{1}{n!} \sum_{\sigma} (\sharp \text{ fixed points of } \sigma \text{ on } T)^2 = \|\chi^{std}\|^2.$$

□

As we have seen, this implies that ρ^{std} is a sum of two distinct irreducible representations (Remark 12.2.2). But, we also know that

$$\rho^{std} = \mathbb{1} \oplus \rho^{std,0}.$$

Therefore, we conclude that $\rho^{std,0}$ is irreducible. This argument is much more elegant, I think, than the proof we previously gave for the irreducibility of $\rho^{std,0}$.[22]

12.4. **The character group G^*.** Recall from §9.1 the set

$$G^* = \mathrm{Hom}(G, \mathbb{C}^\times),$$

which is group under

$$(\phi_1 \cdot \phi_2)(g) = \phi_1(g) \cdot \phi_2(g).$$

For a 1-dimensional representation there is no difference between the representation and its character. The following properties are not hard to prove and the details are left as an exercise:

(1) We have a canonical isomorphism

$$(G_1 \times \cdots \times G_a)^* = G_1^* \times \cdots \times G_a^*.$$

It is given by

$$f \mapsto (f|_{G_1}, \dots, f|_{G_a}),$$

where we identify G_i with $\{1\} \times \cdots \times G_i \times \cdots \times \{1\}$. The inverse isomorphism is given by

$$(f_1, \dots, f_a) \mapsto f_1 \times \cdots \times f_a,$$

where

$$(f_1 \times \cdots \times f_a)(g_1, \dots, g_a) = f_1(g_1) f_2(g_2) \cdots f_a(g_a).$$

(2) We have a canonical isomorphism $G^* \cong (G^{ab})^*$.

[22] Conversely, if we use that $\rho^{std,0}$ is irreducible, the lemma is an immediate consequence of the isomorphism $\rho^{std} = \mathbb{1} \oplus \rho^{std,0}$.

(3) We have a canonical isomorphism

$$(\mathbb{Z}/n\mathbb{Z})^* \cong \mu_n,$$

where $\mu_n = \{e^{j\cdot 2\pi i/n} : j = 0, 1, \dots, n-1\}$ is the multiplicative group of n-th roots of unity in $\mathbb{C}$. (Don't confuse $(\mathbb{Z}/n\mathbb{Z})^*$ with $(\mathbb{Z}/n\mathbb{Z})^\times$.) The isomorphism is given by

$$f \mapsto f(1) \in \mu_n,$$

and

$$\zeta \mapsto f \in (\mathbb{Z}/n\mathbb{Z})^*, \qquad f(a) := \zeta^a.$$

As every finite abelian group is isomorphic to a product of groups of the form $\mathbb{Z}/n\mathbb{Z}$, we have a method to determine G^* for any finite group G:

- Calculate G^{ab}. Any $f\colon G^{ab} \to \mathbb{C}^\times$ induces an element of G^*, i.e., $f \circ \pi$, where $\pi\colon G \to G^{ab}$ is the canonical homomorphism. All multiplicative characters of G arise this way.
- Write $G^{ab} \cong \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_a\mathbb{Z}$. Use the isomorphism $(\mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_a\mathbb{Z})^* \cong (\mathbb{Z}/n_1\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/n_a\mathbb{Z})^*$.
- Use the identification $(\mathbb{Z}/n\mathbb{Z})^* \cong \mu_n$ to fined the multiplicative characters of $\mathbb{Z}/n\mathbb{Z}$.

In particular, we conclude that if G is a finite abelian group then

$$\sharp G = \sharp G^* = h(G).$$

Even better, we can conclude the following corollary of unique decomposition for representations.

Corollary 12.4.1. *Every irreducible representation of an abelian group G is 1-dimensional and there are $\sharp G$ of them. Every n-dimensional representation of G is isomorphic to a representation of the form*

$$\rho\colon G \to \mathrm{GL}_n(\mathbb{C}), \qquad g \mapsto \begin{pmatrix} \alpha_1(g) & & \\ & \ddots & \\ & & \alpha_n(g) \end{pmatrix},$$

for some $\alpha_i \in G^*$.

12.5. **Twisting.** Let (ρ, V) be a representation of G and let $\alpha\colon G \to \mathbb{C}^\times$ be a 1-dimensional representation of G. Then $\mathrm{Hom}((\alpha, \mathbb{C}), (\rho, V))$ is a representation of G of the same dimension and its character, by Theorem 11.1.6, is just

$$\chi_\rho \cdot \bar{\alpha}.$$

As $\bar{\alpha}\colon G \to \mathbb{C}^\times$ is likewise a 1-dimensional representation, we conclude that also $\chi_\rho \cdot \alpha$ is a character. We call the operation $\chi_\rho \mapsto \chi_\rho \cdot \alpha$ **twisting** the representation ρ by the character α. We proved the first part of the following proposition.

Proposition 12.5.1. *For any character χ of G and any 1-dimensional character α of G, also $\chi \cdot \alpha$ is a character. Moreover, if χ is irreducible, so is $\chi \cdot \alpha$.*

Proof. It is not hard to give a direct simple proof of the second part, but let us use characters instead. We have

$$\|\chi\alpha\|^2 = \frac{1}{\sharp G}\sum_g \chi(g)\alpha(g)\bar{\alpha}(g)\bar{\chi}(g).$$

However, because α is 1-dimensional, $\alpha(g)$ is a root of unity and we find

$$\|\chi\alpha\|^2 = \frac{1}{\sharp G}\sum_g \chi(g)\bar{\chi}(g) = \|\chi\|^2 = 1.$$

Thus, by Corollary 12.2.1, χ is irreducible. □

Remark 12.5.2. It is possible that $\chi \cdot \alpha = \chi$ even if $\alpha \neq \mathbb{1}$. In fact, this happens quite often, for example in cases where G has a unique irreducible representation of a given dimension. Nevertheless, in general, twisting by 1-dimensional characters is a very useful method to get new irreducible representations from known ones.

13. Character of induction and Frobenius reciprocity

13.1. **The character of an induced representation.** We return now to the subject of induced representations. Let $H < G$ be finite groups and let (ρ, V) be a representation of H. We defined a representation of G,

$$\mathrm{Ind}_H^G \rho = \mathbb{C}[G] \otimes_{\mathbb{C}[H]} V.$$

We now analyze this representation more closely, calculate its character and prove a remarkable theorem of Frobenius connecting induction and restriction of representations. For another model for the induced representation see Remark 21.2.1.

Denote the character of ρ by χ. We want to the determine the character of $\mathrm{Ind}_H^G\rho$ in terms of χ. We will denote it $\mathrm{Ind}_H^G\chi$.

Choose a set of coset representations for H in G so that $G = \coprod_{i=1}^d g_i H$. Then $\mathbb{C}[G] = \oplus_{i=1}^d g_i\mathbb{C}[H]$ as a right $\mathbb{C}[H]$-module and, since $\mathbb{C}[H] \otimes_{\mathbb{C}[H]} V \cong V$, we have a decomposition

$$\mathbb{C}[G] \otimes_{\mathbb{C}[H]} V \cong (\oplus_{i=1}^d g_i\mathbb{C}[H]) \otimes_{\mathbb{C}[H]} V \cong \oplus_{i=1}^d g_i(\mathbb{C}[H] \otimes_{\mathbb{C}[H]} V) \cong \oplus_{i=1}^d g_i \otimes V.$$

How does G act? Given $g \in G$, we have for any coset representative g_i

$$gg_i = g_j h,$$

for some unique coset representative g_j and unique $h \in H$. Of course, j and $h \in H$ depend on g and i. In fact, if we write $j = j(g,i), h = h(g,i)$, then for any fixed g the map $i \mapsto j(g,i)$ is a permutation of $\{1,\dots,d\}$ and $h = g_j^{-1}gg_i$. We then have for $v \in V$

$$g \cdot g_i \otimes v = gg_i \otimes v = g_j h \otimes v = g_j \otimes \rho(h)(v).$$

Thus, the action of g can be imagined as a block matrix of size $d \times d$, where the $(i, j(g,i))$-block is the matrix $\rho(h(g,i))$. Only the blocks on the diagonal contribute to the trace, and we get a diagonal block at the place i precisely when

$$gg_i = g_i h(g,i),$$

and then $h(g,i) = g_i^{-1}gg_i$. The corresponding block is then $\rho(g_i^{-1}gg_i)$ that has trace $\chi(g_i^{-1}gg_i)$. (N.B., we *cannot* say $\chi(g_i^{-1}gg_i) = \chi(g)$ as $g_i \notin H$ in general.) To sum up this discussion, we have proven the important formula:

$$\mathrm{Ind}_H^G\chi(g) = \sum_{\{i : g_i^{-1}gg_i \in H\}} \chi(g_i^{-1}gg_i).$$

Note that changing g_i to $g_i h$ for $h \in H$ doesn't change the condition $g_i^{-1}gg_i \in H$, nor the value $\chi(g_i^{-1}gg_i)$. Thus, we can reformulate our calculation as follows.

Theorem 13.1.1. *The induced character is given by the formula*

$$\mathrm{Ind}_H^G\chi(g) = \frac{1}{|H|}\sum_{\{b\in G: b^{-1}gb\in H\}}\chi(b^{-1}gb). \tag{17}$$

A common notation is to write

$$\dot\chi$$

for the extension by zero of the character χ to G. Namely, $\dot\chi(x) = \chi(x)$ if $x \in H$ and 0 otherwise. We can then write the character of the induced representation as

$$\mathrm{Ind}_H^G\chi(g) = \tfrac{1}{|H|}\textstyle\sum_{b\in G}\dot\chi(b^{-1}gb). \tag{18}$$

Corollary 13.1.2. *Suppose that $H\lhd G$ then $b^{-1}gb \in H$ iff $g \in H$ and so we find*

$$\mathrm{Ind}_H^G\chi(g) = \begin{cases} 0 & g\notin H \\ \frac{1}{|H|}\sum_{b\in G}\chi(b^{-1}gb) & g\in H. \end{cases}$$

Example 13.1.3. From the formula we find that the character χ^{reg} of the regular representation $\mathbb{C}[G] = \mathrm{Ind}_{\{1\}}^G\mathbb{C}$ satisfies $\chi^{reg}(g) = 0$ if $g \neq 1$ and $\chi^{reg}(1) = |G|$. It is of course easy to deduce that from the definition of the representation itself.

Example 13.1.4. Let $\zeta_1 = 1, \zeta_2 = e^{2\pi i/3}, \zeta_3 = e^{4\pi i/3}$. The three 1-dimensional representations ρ_i of A_3 are determined by

$$\rho_i((123)) = \zeta_i.$$

Let $\chi_i = \mathrm{Ind}_{A_3}^{S_3}\rho_i$; we can easily calculate it using Corollary 13.1.2. The following table gives the values of these characters on the three conjugacy classes of S_3.

	1	(12)	(123)
χ_1	2	0	2
χ_2	2	0	-1
χ_3	2	0	-1

As two representations with the same character are isomorphic, we find here an example of two non-isomorphic representations of a subgroup, that is ρ_2, ρ_3, whose induced representations are isomorphic. We can also calculate the norm and find that χ_2, χ_3 are irreducible, while χ_1 is not. It is isomorphic to the direct sum of the trivial representation $\mathbb{1}$ and the one-dimensional sign representation $\mathrm{sgn} : S_3 \to \{\pm 1\}$.

13.2. **Frobenius reciprocity.** Frobenius reciprocity in its precise form is a very useful tool to analyze induced representations. Its importance rests on the fact that induced representations are a very powerful method to arrive at the irreducible representations of a group G starting from its subgroups. We have already seen a form of Frobenius reciprocity when studying the $(\otimes, \mathrm{Hom})$ adjoint pair.

Given a group G we will sometimes add the subscript G in $\langle\alpha,\beta\rangle_G$ when denoting the inner product of its characters.

Theorem 13.2.1. (Frobenius Reciprocity) *Let H be a subgroup of a finite group G and let V be a representation of H with character ρ and W a representation of G with character σ. Then*

$$\langle \mathrm{Ind}_H^G \rho, \sigma \rangle_G = \langle \rho, \mathrm{Res}_H^G \sigma \rangle_H.$$

Proof. We just calculate!

$$\begin{aligned}
\langle \mathrm{Ind}_H^G \rho, \sigma \rangle_G &= \frac{1}{|G|} \sum_{g \in G} \mathrm{Ind}_H^G \rho(g) \cdot \overline{\sigma(g)} \\
&= \frac{1}{|G|} \sum_{g \in G} \left(\frac{1}{|H|} \sum_{h \in G} \dot{\rho}(h^{-1} g h) \right) \cdot \overline{\sigma(g)} \\
&= \frac{1}{|H| \cdot |G|} \sum_{t \in H} \sum_{\{g,h \in G : h^{-1} g h = t\}} \rho(t) \cdot \overline{\sigma(hth^{-1})} \\
&= \frac{1}{|H|} \sum_{t \in H} \rho(t) \cdot \overline{\sigma(t)} \\
&= \langle \rho, \sigma \rangle_H.
\end{aligned}$$

We used that in $\{g, h \in G : h^{-1} g h = t\}$ actually $g = hth^{-1}$ and so is uniquely determined by h that can be arbitrary, and that for $h, t \in G$ we have $\sigma(hth^{-1}) = \sigma(t)$ because σ is a representation of G. □

We note an immediate conclusion: if ρ and σ are irreducible representations of H and G, respectively, then the multiplicity to which σ appears in the induced representation $\mathrm{Ind}_H^G \rho$ is equal to the multiplicity to which ρ appears in σ restricted to the subgroup H. Applying Frobenius reciprocity, one may deduce the following (details left as exercise).

Corollary 13.2.2. *Let $H \lhd G$ and let (σ, W) be an irreducible representation of H. Then $\mathrm{Ind}_H^G \sigma$ is irreducible if and only if for all $g \in G \setminus H$ the representation*

$$\sigma^g : H \to \mathrm{GL}(W), \quad \sigma^g(h) = \sigma(g^{-1} h g),$$

is not isomorphic to σ. In particular, we must have $\mathrm{Cent}_G(H) \subseteq H$.

13.3. **Representations of D_n.** We apply here the technique of induction to find the irreducible representations of the dihedral group D_n. There are minor variations between the case n even and n odd. We consider here the case of $n \geq 3$ odd, leaving the case of n even as an exercise.

We use the usual presentation of $D_n = \langle x, y : y^2 = x^n = yxyx = 1 \rangle$. First, we note that $yxy^{-1}x^{-1} = x^{-2}$ and so, as n is odd, the commutator subgroup of D_n contains $\langle x \rangle$. Since $D_n / \langle x \rangle$ is abelian, it follows that $D_n' = \langle x \rangle$ and $D_n^{ab} \cong \mathbb{Z}/2\mathbb{Z}$. Thus, D_n has two 1-dimensional characters: one is the trivial character $\mathbb{1}$, and the other character ψ is determined by $\psi(x) = 1, \psi(y) = -1$.

Our analysis below will reveal that any other irreducible representation of D_n is 2-dimensional, induced from a character of $H := \langle x \rangle$ and that there are $(n-1)/2$ of them. Note that

$$\frac{(n-1)}{2} \times 2^2 + 1 + 1 = 2n,$$

and this identity not only confirms the dimension formula $\sharp G = \sum d_i^2$ but also shows that it if we have found $(n-1)/2$ non-isomorphic irreducible 2-dimensional representations of D_n then, together with the 1-dimensional representations, we have found them all.

Let $\chi\colon H\to\mathbb{C}^\times$ be a multiplicative non-trivial character of H. Then $\chi(x)=\zeta$, where ζ is some non-trivial n-th root of unity, and $\chi(x^i)=\zeta^i$. We claim that the 2-dimensional representation $\mathrm{Ind}_H^{D_n}\chi$ is irreducible. Let α denote its character. Let $x^i\in H$ then $yx^iy^{-1}=x^{-i}$. Since every element of D_n is either x^a or x^ay for some a, we can easily calculate all the conjugates of x^i. Using Corollary 13.1.2 we find that

$$\alpha(x^iy)=0,\quad \alpha(x^i)=\zeta^i+\zeta^{-i}.$$

To show $\mathrm{Ind}_H^{D_n}\chi$ is irreducible, we can use Corollary 13.2.2. Let $g=x^ay$ then $gx^ig^{-1}=x^{-i}$ and so we only need to verify that the character χ, given by $x^i\mapsto\zeta^i$, is not the character χ^g, given by $x^i\mapsto\zeta^{-i}$. Take $i=1$. We can't have $\zeta=\zeta^{-1}$ because that means $\zeta^2=1$, but n is odd. Thus, $\mathrm{Ind}_H^{D_n}\chi$ is indeed irreducible. We remark that an alternative way to prove that is to calculate that $\|\alpha\|^2=1$, which we leave as an exercise (it uses that ζ^2 is a non-trivial root of unity and that for a non-trivial root of unity γ of order m, $1+\gamma+\gamma^2+\cdots+\gamma^{m-1}=0$).

Finally, our formula for α shows that $\mathrm{Ind}_H^{D_n}\chi\cong\mathrm{Ind}_H^{D_n}\chi'$ if and only $\chi(x)=\chi'(x)^{\pm1}$. Thus, if we denote χ_a the character of H such that $\chi_a(x)=e^{a\cdot2\pi i/n}$, the characters χ_a and χ_{n-a} give the same induced representation, and the representations

$$\mathrm{Ind}_H^{D_n}\chi_a,\quad a=1,\dots,(n-1)/2,$$

are representatives for the isomorphism classes of all the irreducible representations of dimension greater than 1.

14. Character tables

The character table of a group G is one of the best ways to get insight into the structure of G and its action on vector spaces. There are whole books written on this subject.[23] In this section we will study various properties of the character table. Our treatment is by no means complete: there are additional properties we will not even mention, and there are properties that we will mention but not prove.

The **character table** of G has rows for every irreducible representation of G, and columns for every conjugacy class of G. We reserve the first row for the character $\mathbb{1}$ and the first column for the conjugacy class of the identity (often we will write a representative element for each conjugacy class, and indicate below the conjugacy class how many elements it contains). The table entry corresponding to a character χ and a conjugacy class c is just $\chi(c)$. By that we mean $\chi(x)$ for any $x\in c$; the choice of x doesn't influence the value $\chi(x)$. So, for example, the character table of S_3 is the following:

	1	(12)	(123)
	1	3	2
$\chi_1=\mathbb{1}$	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

Table 1: Character table of S_3

We see the three representatives $1,(12),(123)$ of the distinct conjugacy classes of S_3 and the sizes of their conjugacy classes are indicated by $1,3,2$. We see 3 irreducible characters. The first one is the trivial character $\mathbb{1}$, the second is the sign homomorphism $\mathrm{sgn}\colon S_3\to\mathbb{C}^\times$, and the third is the character $\chi^{std,0}$.

[23] For example: I. Martin Isaacs, *"Character Theory of Finite Groups"*, Dover 1994.

We will usually use the notation χ_i for the rows and c_i for the columns. We use the notation introduced before: χ_i is the character of the irreducible representation ρ_i that has dimension d_i.

14.1. **First properties of the character table.**

Theorem 14.1.1. *The character table of G has the following properties:*

1. *The number of rows equals to the number of columns.*
2. *The sum of the squares of the entries of the first column is the cardinality of the group.*
3. *The number of rows with 1 in the first column is equal to $\sharp G^{ab}$.*
4. *Every entry in the first column is an integer dividing $\sharp G$.*
5. *The "weighted" inner-product of distinct rows is 0. The weighted self-product of a row is equal to $\sharp G$ (here the weights are the cardinality of conjugacy classes).*
6. *The "weighted" sum of the rows is the vector $(\sharp G, 0, \ldots, 0)$ (here the weights are the dimensions of the representations).*

The proof consists of references to theorems we proved, or will prove shortly.

Proof. (1) This is the statement that the number of irreducible characters h is actually equal to $h(G)$. We mentioned this before and we will prove it in Theorem 15.1.1 below.

(2) This is Corollary 12.1.2: $\sharp G = \sum_{i=1}^{h} d_i^2$

(3) This states that the irreducible characters of dimension 1 are 1-dimensional characters $G \to \mathbb{C}^\times$, and $\sharp G^* = \sharp (G^{ab})^*$ (Lemma 9.1.1).

(4) This is a theorem we will prove later (Theorem 18.1.1).

(5) This is just orthogonality of characters (Theorem 11.4.1). If we use the fact that characters are class functions, we may write

$$\langle \chi_i, \chi_j \rangle = \frac{1}{\sharp G} \sum_{g \in G} \chi_i(g) \bar{\chi}_j(g) = \frac{1}{\sharp G} \sum_{i=1}^{h} |c_i| \cdot \chi_i(c_i) \bar{\chi}_j(c_i).$$

We find that if $i \neq j$ the weighted inner-product of the rows, $\sum_{i=1}^{h} |c_i| \chi_i(c_i) \bar{\chi}_j(c_i)$, is equal to 0, and if $i = j$ it is equal to $\sharp G$.

(6) This is just a restatement of the decomposition of the regular representation: $\chi^{reg} = \sum_{i=1}^{h} d_i \chi_i$ (Proposition 12.1.1). □

14.2. **Examples of character tables.**

14.2.1. *Character table of $\mathbb{Z}/n\mathbb{Z}$.* Recall that every irreducible representation of an abelian group is a multiplicative character and that we have

$$(\mathbb{Z}/n\mathbb{Z})^* \cong \mu_n.$$

We usually denote the corresponding characters $\rho_0, \ldots, \rho_{n-1}$ in this case, because if we let $\zeta = e^{2\pi i/n}$ then we have

$$\rho_i(a) = \zeta^{ai}.$$

(This notation is slightly in conflict with the usual convention of denoting the irreducible characters of a group G by $\chi_1, \ldots, \chi_h$.) We find the following table

	0	1	2		$n-1$
$\rho_0 = \mathbb{1}$	1	1	1	...	1
ρ_1	1	ζ	ζ^2	...	ζ^{n-1}
ρ_2	1	ζ^2	ζ^4	...	$\zeta^{2(n-1)}$
$\vdots$				$\vdots$	
ρ_{n-1}	1	ζ^{n-1}	$\zeta^{2(n-1)}$	...	$\zeta^{(n-1)^2}$

Table 2: Character table of $\mathbb{Z}/n\mathbb{Z}$

Note that property (6) in Theorem 14.1.1 gives us the very useful identity: For a root of unity ζ of order n, we have $\sum_{i=0}^{n-1} \zeta^{ai} = 0$ for every $a \not\equiv 0(n)$.

14.2.2. *Character tables of* $(\mathbb{Z}/2\mathbb{Z})^2$, $\mathbb{Z}/3\mathbb{Z}\times\mathbb{Z}/5\mathbb{Z}$, $(\mathbb{Z}/3\mathbb{Z})^2$. "Multiplying" two copies of the character table of $\mathbb{Z}/2\mathbb{Z}$ we find

	0	1
$\mathbb{1}$	1	1
ρ_1	1	-1

$\times$

	0	1
$\mathbb{1}$	1	1
ρ_1	1	-1

$=$

	(0,0)	(1,0)	(0, 1)	(1,1)
$\mathbb{1}\times\mathbb{1}$	1	1	1	1
$\mathbb{1}\times\rho_1$	1	1	-1	-1
$\rho_1\times\mathbb{1}$	1	-1	1	-1
$\rho_1\times\rho_1$	1	-1	-1	1

Table 3: Character table of $(\mathbb{Z}/2\mathbb{Z})^2$

Similarly, for any abelian group $G \cong \mathbb{Z}/n_1\mathbb{Z}\times\cdots\times\mathbb{Z}/n_a\mathbb{Z}$ we can "multiply" the character tables for each $\mathbb{Z}/n_i\mathbb{Z}$ to find the character table of G. This rests on our discussion that showed

$$G^* \cong (\mathbb{Z}/n_1\mathbb{Z})^*\times\cdots\times(\mathbb{Z}/n_a\mathbb{Z})^* \cong \mu_{n_1}\times\cdots\times\mu_{n_a},$$

and the concrete description of the character table of $\mathbb{Z}/n\mathbb{Z}$.

It is not efficient to use this method for $G = \mathbb{Z}/3\mathbb{Z}\times\mathbb{Z}/5\mathbb{Z}$ because by CRT we have $G \cong \mathbb{Z}/15\mathbb{Z}$ which is a cyclic group for which we already have a nice description. But, for example, for the case $G = (\mathbb{Z}/3\mathbb{Z})^2$ it is useful, and we find the following 9×9 table ($\omega = e^{2\pi i/3}$):

	0	1	2
$\mathbb{1}$	1	1	1
ρ_1	1	ω	ω^2
ρ_2	1	ω^2	ω

$\times$

	0	1	2
$\mathbb{1}$	1	1	1
ρ_1	1	ω	ω^2
ρ_2	1	ω^2	ω

$=$

	(0,0)	...	(1, 2)	...	(a,b)
$\mathbb{1}\times\mathbb{1}$	1		1		1
$\vdots$					
$\rho_1\times\rho_2$	1		ω^2		ω^{a+2b}
$\vdots$		$\vdots$		$\vdots$	
$\rho_i\times\rho_j$	1		ω^{i+2j}		ω^{ai+bj}
$\vdots$		$\vdots$		$\vdots$	

Table 4: Character table of $(\mathbb{Z}/3\mathbb{Z})^2$

14.2.3. *Character table of S_3.* We have $h(S_3) = p(3) = 3$ and so there are 3 conjugacy classes and we take as representatives $1, (12), (123)$. Their sizes are $1, 3, 2$, respectively. We have

$$S_3^{ab} = S_3/A_3 \cong \mathbb{Z}/2\mathbb{Z},$$

and, in fact, we know two 1-dimensional characters: $\mathbb{1}$ and $\mathbf{sgn}$. As we must have

$$\sharp S_3 = 6 = 1^1 + 1^1 + x^2,$$

we conclude that the remaining irreducible representation of S_3 is 2-dimensional. We happen to know such a representation, namely, $\rho^{std,0}$ and its character $\chi^{std,0}$ whose value on a permutation σ is the number of fixed points of σ minus 1. Or, using the isomorphism $S_3 \cong D_3$ and our description of the characters of dihedral groups, we can find that representation as induced from a non-trivial multiplicative character of A_3. We therefore find the following table:

	1	(12)	(123)
	1	3	2
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

Table 5: Character table of S_3

Remark though that we didn't really need to use our "lucky break" of knowing before-hand an irreducible 2-dimensional representation. We could have *solved* for the remaining character:

$$\chi_3 = \frac{1}{2}(\chi^{reg} - \chi_1 - \chi_2)$$

(Theorem 14.1.1).

14.2.4. *Character table of D_4.* It requires some calculations but one find that

$$D_4' = \{1, x^2\}, \quad D_4^{ab} = \{1, \bar{x}, \bar{y}, \bar{x}\bar{y}\},$$

and that every element of D_4^{ab} has order 2. Thus,

$$D_4^{ab} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \qquad \bar{x} \mapsto (1,0), \bar{y} \mapsto (0,1).$$

We also calculate "by hand" the conjugacy classes and find that they are given by

$$c_1 = \{1\},\ c_2 = \{x, x^{-1}\},\ c_3 = \{x^2\},\ c_4 = \{y, yx^2\},\ c_5 = \{yx, yx^{-1}\}.$$

There isn't a really quick way to do that, but one can note that since $\langle x \rangle$ is a normal subgroup, conjugacy classes are either contained in it, or disjoint from it. At any rate, we now know that D_4 has four 1-dimensional representations, "lifted" from $(\mathbb{Z}/2\mathbb{Z})^2$. That is, if χ is an irreducible character of $(\mathbb{Z}/2\mathbb{Z})^2$ and f is the composition $D_4 \to D_4^{ab} \to (\mathbb{Z}/2\mathbb{Z})^2$ then $\chi \circ f$ is an 1-dimensional character of D_4.

In addition, D_4 has one more irreducible representation and its dimension x satisfies

$$8 = \sharp D_4 = 1^2 + 1^2 + 1^2 + 1^2 + x^2.$$

It follows that we are missing a 2-dimensional representation. Note that we can solve for the missing character, say χ, using the result on the sum of the rows of the character table, but it is also natural to wonder whether the missing representation is provided by the action of D_4 on the plane (the action inducing the action of D_4

on the square). In this representation ρ^{pl}, the action of the representatives for conjugacy classes is given as follows:

$$1 = \left(\begin{smallmatrix} 1 & \\ & 1 \end{smallmatrix}\right),\ x = \left(\begin{smallmatrix} & 1 \\ -1 & \end{smallmatrix}\right),\ y = \left(\begin{smallmatrix} -1 & \\ & 1 \end{smallmatrix}\right),\ x^2 = \left(\begin{smallmatrix} -1 & \\ & -1 \end{smallmatrix}\right),\ yx = \left(\begin{smallmatrix} & -1 \\ -1 & \end{smallmatrix}\right).$$

We can now write the character table of D_4. The last row is $\chi^{pl} = \chi_{\rho^{pl}}$, which is indeed irreducible because $\|\chi^{pl}\| = 1$. (Yet another method is to construct this representation as induced from a one-dimensional representation of the subgroup $\langle x \rangle$.)

	1 1	x 2	y 2	xy 2	x^2 1
$\mathbb{1}$	1	1	1	1	1
$\rho_1 \times \mathbb{1}$	1	-1	1	-1	1
$\mathbb{1} \times \rho_1$	1	1	-1	-1	1
$\rho_1 \times \rho_1$	1	-1	-1	1	1
χ^{pl}	2	0	0	0	-2

Table 6: Character table of D_4

Here is an application. The composition ρ defined by

$$D_4 \longrightarrow S_4 \xrightarrow{\rho^{std}} \mathrm{GL}_4(\mathbb{C}),$$

(where the first arrow is the natural inclusion of D_4 into S_4, $x \mapsto (1234), y \mapsto (24)$) is a 4-dimensional representation of D_4. It is a bit hard to understand this action. Indeed, in terms of matrices

$$x = \begin{pmatrix} 0&0&0&1 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \end{pmatrix},\ y = \begin{pmatrix} 1&0&0&0 \\ 0&0&0&1 \\ 0&0&1&0 \\ 0&1&0&0 \end{pmatrix},$$

and it is not easy to understand what is the overall action of elements of the group. However, we can decompose ρ into irreducible representations. A calculation gives

$$\langle \chi_\rho, \mathbb{1} \rangle = 1 \quad , \langle \chi_\rho, \rho_1 \times \mathbb{1} \rangle = 1, \quad \langle \chi_\rho, \chi^{pl} \rangle = 1.$$

This tells us that

$$\rho \cong \mathbb{1} \oplus (\rho_1 \times \mathbb{1}) \oplus \rho^{pl}.$$

That means that there is another basis for $\mathbb{C}^4$ in which the representation has the form

$$x = \begin{pmatrix} 1&0&0&0 \\ 0&-1&0&0 \\ 0&0&0&1 \\ 0&0&-1&0 \end{pmatrix},\ y = \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&-1&0 \\ 0&0&0&1 \end{pmatrix}.$$

And a general element g of D_4 will act by a matrix of the form

$$\begin{pmatrix} 1&0&0 \\ 0&\pm 1&0 \\ 0&0&\rho^{pl}(g) \end{pmatrix}.$$

It is now much easier to understand the action of D_4.

14.2.5. *Character table of S_4.* Here is a general principle. Let $f\colon A \to B$ be a homomorphism of groups. Let $\rho\colon B \to \mathrm{GL}(V)$ be a representation of B. Then $\rho \circ f$ is a representation of A and its character is simply

$$\chi_{\rho\circ f} = \chi_\rho \circ f\colon A \to \mathbb{C}.$$

In fact, we have used it several times before in the situation $G \to G^{ab} \to \mathbb{C}^\times$ to lift 1-dimensional characters of G^{ab} to G.

Now, if f is *surjective* and ρ is irreducible then also $\rho \circ f$ is irreducible. Indeed, suppose that $U \subseteq V$ is a subrepresentation of $\rho \circ f$. That is, for all $a \in A$ we have $\rho(f(a))(U) \subseteq U$. Then, as f is surjective, it follows that for all $b \in B$ we have $\rho(b)(U) \subseteq U$. It follows that U is a subrepresentation of ρ and so $U = 0$ or V.

Let us use this for the surjective homomorphism $f\colon S_4 \to S_3$, whose kernel is K, the Kline group. (It can be constructed by letting S_4 act by conjugation on the set $K - \{1\} = \{(12)(34), (13)(24), (14)(23)\}$.) Using it, we can lift the characters of S_3 to S_4, and so we easily find the first 3 rows of the character table of S_4. (The conjugacy classes of S_n correspond to the cycle type of permutations and that gives us the columns' labels.) As there are 5 conjugacy classes, there are two additional irreducible representations. We know one of them, $\rho^{std,0}$, and we get the last row as the twist $\rho^{std,0} \cdot \mathrm{sgn}$ (or by solving the equation where the sum of the rows with multiplicities is equal to the vector $(24, 0, 0, 0, 0)$). We find the following table.

	1	(12)	(123)	(1234)	$(12)(34)$
	1	6	8	6	3
$\mathbb{1}$	1	1	1	1	1
sgn	1	-1	1	-1	1
$\chi_3 \circ f$	2	0	-1	0	2
$\chi^{std,0}$	3	1	0	-1	-1
$\chi^{std,0} \cdot \mathrm{sgn}$	3	-1	0	1	-1

Table 7: Character table of S_4

14.2.6. *Character table of A_4.* The representatives for the conjugacy classes of A_4 are given by $1, (12)(34), (123), (132)$. There are therefore 4 irreducible representations. As A_4/K is of order 3, it follows that $A_4/K \cong \mathbb{Z}/3\mathbb{Z}$ and that $K \supseteq A_4'$. As A_4 is not abelian, $A_4' \neq \{1\}$ and so contains some element of cycle type $(2,2)$. But those form a single conjugacy class and A_4' is normal. It follows that $A_4' = K$.

We conclude that there are 4 irreducible representations, of which 3 are 1-dimensional, and the last is 3-dimensional (as $\sharp A_4 = 1^2 + 1^2 + 1^2 + x^2$ only allows $x = 3$). Using the result about the sum of rows we find the following character table:

	1	(123)	(132)	$(12)(34)$
	1	4	4	3
$\mathbb{1}$	1	1	1	1
χ_1	1	ω	ω^2	1
χ_2	1	ω^2	ω	1
χ	3	0	0	-1

Table 8: Character table of A_4

It turns out that the last character is just $\chi^{std,0}|_{A_4}$. This is no coincidence. One can prove that for $n \geq 4$ the representation $\rho^{std,0}|_{A_n}$ is an irreducible representation of A_n.

14.3. **Orthogonality of columns.** We show here that the columns of the character table enjoy an orthogonality property. We begin with some renormalization device to make the argument more transparent.

For every character χ of G (or even for every class function f), we define a vector $v_\chi \in \mathbb{C}^h$, where $h = h(G)$ is the number of conjugacy classes of G. Let $c_1, \ldots, c_h$ be the conjugacy classes of G, and let

$$v_\chi = (\sqrt{\frac{\sharp c_1}{\sharp G}} \cdot \chi(c_1), \ldots, \sqrt{\frac{\sharp c_h}{\sharp G}} \cdot \chi(c_h))$$

The point of this construction is that for every two characters χ, ψ (or even any two class functions) we have

$$\langle \chi, \psi \rangle = \langle v_\chi, v_\psi \rangle,$$

where the inner-product on the left is the inner product of class-functions, and the inner-product on the right is the usual inner-product in $\mathbb{C}^h$. In fact, we have already noticed something very similar – see the proof of part (5) of Theorem 14.1.1.

Let $\chi_1, \ldots, \chi_h$ denote the irreducible characters of G. It follows that the rows of the following matrix are orthonormal:

$$\begin{pmatrix} \hline v_{\chi_1} \\ \hline v_{\chi_2} \\ \hline \vdots \\ \hline v_{\chi_h} \\ \hline \end{pmatrix}.$$

But this implies that the columns of the same matrix are an orthonormal set too. Namely, for any two conjugacy classses c_a, c_b we get that

$$\sum_{i=1}^{h} \sqrt{\frac{\sharp c_a}{\sharp G}} \sqrt{\frac{\sharp c_b}{\sharp G}} \cdot \chi_i(c_a)\bar{\chi}_i(c_b) = \delta_{ab}.$$

Note that $\sharp(G)/\sharp(c_a) = \sharp Cent(x)$ for any $x \in C_a$. Therefore, we conclude the following.

Proposition 14.3.1 (Orthogonality of columns)**.** *We have the following orthogonality properties of the columns of the character table.*

(1) If $c_a \neq c_b$ are conjugacy classes then the product of the c_a column with the c_b column is 0. To be precise:

$$\sum_{i=1}^{h} \chi_i(c_a)\bar{\chi}_i(c_b) = 0.$$

(2) For every conjugacy class c_a the norm of the c_a column is the cardinality of its centralizer. That is,

$$\sum_{i=1}^{h} |\chi_i(c_a)|^2 = \sharp Cent(x), \quad x \in c_a.$$

It follows that we can use the entries of the character table, more specifically we can use the second part of the proposition, to figure out the size of conjugacy classes. We record it as a corollary.

Corollary 14.3.2. *The character table determines the size of the conjugacy classes.*

15. Irreducible characters form a basis for Class(G)

In this section we fill a gap and prove that the irreducible characters of a group G form a basis for Class(G). Nothing prevented us from proving it sooner; it just seemed more useful to see some examples before developing the theory further.

15.1. **Irreducible characters form a basis.**

Theorem 15.1.1. *Let G be a group and let $\chi_1, \ldots, \chi_h$ be its irreducible characters. Then*

$$\{\chi_1, \ldots, \chi_h\}$$

is an orthonormal basis for Class(G).

Proof. We begin with a lemma that constructs endomorphisms of representations.

Lemma 15.1.2. *Let (ρ, V) be a representation of G and let α a class function. Then the linear operator*

$$T = T_\rho = \sum_{g \in G} \alpha(g)\rho(g) \in \mathrm{End}_G(V).$$

Proof. The fact that T is a linear operator is clear, because $\alpha(g)$ are scalars and T is the sum of the linear operators $\alpha(g)\rho(g)$. The point is that it commutes with ρ. We have

$$\rho(h) \circ T \circ \rho(h)^{-1} = \sum_{g \in G} \alpha(g)\rho(hgh^{-1}) = \sum_{g \in G} \alpha(hgh^{-1})\rho(hgh^{-1}).$$

The last equality is true because α is a class function. Now, $g \mapsto hgh^{-1}$ is a bijection of G (even an automorphism) and hence

$$\rho(h) \circ T \circ \rho(h)^{-1} = \sum_{g \in G} \alpha(hgh^{-1})\rho(hgh^{-1}) = \sum_{g \in G} \alpha(g)\rho(g) = T.$$

□

We know already that $\{\chi_1, \ldots, \chi_h\}$ are an orthonormal set. To prove they form a basis we need only show for $\beta \in \mathrm{Class}(G)$,

$$\langle \chi_i, \beta \rangle = 0, \forall i \implies \beta \equiv 0.$$

Let $\alpha = \bar{\beta}$. It will of course be enough to prove $\alpha \equiv 0$.

Let (ρ, V) be an irreducible representation. We claim the the operator

$$T_\rho := \sum_{g \in G} \alpha(g)\rho(g) \in \mathrm{End}_G((\rho, V))$$

is actually the zero operator. By Schur's Lemma, we have $\mathrm{End}_G((\rho, V)) \cong \mathbb{C}$ under the map $T \mapsto \frac{1}{\dim(V)}\mathrm{Tr}(T)$ (Equation (14)). Now,

$$\mathrm{Tr}(T_\rho) = \sum_{g \in G} \alpha(g)\mathrm{Tr}(\rho(g)) = \sum_{g \in G} \chi_\rho(g)\bar{\beta}(g) = \sharp G \langle \chi_\rho, \beta \rangle = 0.$$

And therefore $T_\rho = \frac{1}{\dim(V)}\mathrm{Tr}(T_\rho) \cdot Id_V = 0$.

Note that the construction

$$\rho \mapsto T_\rho = \sum_{g \in G} \alpha(g)\rho(g)$$

commutes with direct sums. Thus, we may conclude that for *any* representation (ρ, V) of G we have $T_\rho = 0$. In particular this holds of the regular representation. That is, we conclude that $\sum_{g\in G}\alpha(g)\rho^{reg}(g)$ is the zero operator on $\mathbb{C}[G]$. In this case, we must have

$$\sum_{g\in G}\alpha(g)\rho^{reg}(g)(e_1) = 0,$$

where $e_1 \in \{e_g : g \in G\}$ is the basis vector indexed by the identity element of G. However,

$$\sum_{g\in G}\alpha(g)\rho^{reg}(g)(e_1) = \sum_{g\in G}\alpha(g)e_g.$$

As $\{e_g\}$ is a basis, it follows that $\alpha(g) = 0$ for all $g \in G$, as we wanted to show. □

15.2. **More properties of the character table.** We organize together all the properties of the character table we have seen, implicitly or explicitly.

Theorem 15.2.1. *Let G be a group with class number h. Let $\{\chi_i : i = 1,\dots,h\}$ be its irreducible characters, $d_i = \dim(\chi_i) = \chi_i(1)$, and let $\{c_a : a = 1,\dots,h\}$ be the conjugacy classes of G. We assume always that $\chi_1 = \mathbb{1}$ and $c_1 = \{1_g\}$.*

The character table of G has the following properties:

1. *The number of rows equals to the number of columns.*
2. *The sum of the squares of the entries of the first column is the cardinality of the group.*
3. *The number of rows with 1 in the first column is equal to $\sharp G^{ab}$.*
4. *Every entry in the first column is an integer dividing $\sharp G$.*
5. *The "weighted" inner-product of distinct rows is 0. The weighted self-product of a row is equal to $\sharp G$ (here the weights are the cardinality of conjugacy classes).*
6. *The "weighted" sum of the rows is the vector $(\sharp G, 0, \dots, 0)$ (here the weights are the dimensions of the representations).*
7. *For any two columns c_a, c_b we have*

$$\sum_{i=1}^{h}\chi_i(c_a)\bar{\chi}_i(c_b) = 0, \quad a \neq b,$$

and

$$\sum_{i=1}^{h}|\chi_i(c_a)|^2 = |Cent(x)|, \ x \in c_a.$$

8. *If $\chi_i(c_a) = \alpha$ then $\chi_i(c_a^{-1}) = \bar{\alpha}$ where c_a^{-1} is the conjugacy class $\{x^{-1} : x \in c_a\}$. In particular, the set of entries of the character table is closed under complex conjugation.*
9. *If χ_i is 1-dimensional and χ_j is any other irreducible character, then $\chi_i \cdot \chi_j = \chi_k$ for some irreducible character χ_k (possibly equal to χ_j).*
10. *$|\chi_i(g)| \leq \chi_i(1)$, with equality if and only if $\rho_i(g) = \alpha \cdot Id$ for some root of unity α.*
11. *If $c_a \neq c_b$ then there is some character χ_i such that $\chi_i(c_a) \neq \chi_i(c_b)$.*
12. *The weighted sum of the columns, where the i-th column is given weight $|c_i|$, is the vector ${}^t(\sharp G, 0, \dots, 0)$.*

Proof. We have already proved properties (1) - (6) in Theorem 14.1.1 (only that now we have really proved (1)). Property (7) is the orthogonality of columns proven in Proposition 14.3.1.

Property (8) was also mentioned before: we have seen that $\chi_i(x^{-1}) = \overline{\chi_i(x)}$ (Equation 12). Property (9) is of course the twisting operation we have studied in § 12.5. Property (10) follows from the fact that $\chi_i(g)$ is a sum of d_i roots of unity and the absolute value is equal to d_i if and only if they all point in the same direction.

Property (11) follows from the fact that the $\{\chi_i\}$ form a basis for the class functions and so for any given $c_a \neq c_b$ a suitable linear combination of them should have value 1 on c_a and value 0 on c_b. This is only possible if for some i, $\chi_i(c_a) \neq \chi_i(c_b)$.

Property (12) is essentially the orthogonality of χ_1 and χ_i for $i \neq 1$. Indeed, the i-th entry of this sum of columns is

$$\sum_{j=1}^{h} |c_j| \chi_i(c_j) = \sharp G \cdot \langle \chi_i, \chi_1 \rangle = \sharp G \cdot \delta_{i,1}.$$

□

Character tables have even more properties. We mention an additional one, which is a theorem of Burnside, just because it is so easy to state (we will not use it in this course)

Theorem 15.2.2 (Burnside)**.** *If $d_i > 1$ then χ_i takes the value 0 for some conjugacy class.*

16. Using the character table to find normal subgroups

We will now see a beautiful application of character tables for the calculation of all normal subgroups of a group G.

16.1. **Normal subgroups and character kernels.** Let (ρ, V) be any representation of G with character $\chi = \chi_\rho$. Define

$$\mathrm{Ker}(\chi_\rho) := \{g \in G : \chi_\rho(g) = \chi(1)\} = \{g \in G : \chi_\rho(g) = \dim(V)\}.$$

Lemma 16.1.1. *We have*

$$\mathrm{Ker}(\chi_\rho) = \mathrm{Ker}(\rho),$$

and so $\mathrm{Ker}(\chi_\rho)$ is a normal subgroup of G.

Proof. Let $g \in \mathrm{Ker}(\rho)$ then $\rho(g) = \mathrm{Id}_V$. Therefore, $\chi(g) = \mathrm{Tr}(\mathrm{Id}_V) = \dim(V)$ and thus $g \in \mathrm{Ker}(\chi)$.

Conversely, let $g \in \mathrm{Ker}(\chi)$ and $d = \dim(V)$. As $\chi(g)$ is a sum of d roots of unity (which are the eigenvalues, with multiplicity, of $\rho(g)$), the only way this sum can be equal to d is if all these roots of unity are 1. This implies $\rho(g) = \mathrm{Id}_V$ and so $g \in \mathrm{Ker}(\rho)$. □

In particular, if $\chi_1, \ldots, \chi_h$ denote the irreducible characters of G, as per our usual notation, we have the normal subgroups

$$\mathrm{Ker}(\chi_i), \quad i = 1, 2, \ldots, h.$$

Note that these subgroups can all be written as a union of specific conjugacy classes, given the character table of G.

Lemma 16.1.2. *Let χ be a character of a representation (ρ, V) of G. Suppose that*

$$\chi = \sum_{i \in I} a_i \chi_i,$$

for a subset $I \subseteq \{1,2,\dots,h\}$ and positive integers a_i. Then,

$$\mathrm{Ker}(\chi) = \cap_{i\in I}\mathrm{Ker}(\chi_i).$$

Once more, note that this can be calculated effectively from the character table of G.

Proof. We have

$$\chi(1) = \sum_{i\in I} a_i\chi_i(1).$$

If $g \in \mathrm{Ker}(\chi_i)$ for every i, then

$$\chi(g) = \sum_{i\in I} a_i\chi_i(g) = \sum_{i\in I} a_i\chi_i(1) = \chi(1),$$

and so $g \in \mathrm{Ker}(\chi)$. Conversely, if $g \in \mathrm{Ker}(\chi)$ we have

$$\chi(1) = \chi(g) = \sum_{i\in I} a_i\chi_i(g) = \sum_{i\in I} a_i\chi_i(1).$$

Since the a_i are positive integers and $|\chi_i(g)| \leq \chi_i(1)$, the only way the last equality can hold is if $\chi_i(g) = \chi_i(1)$ for every $i \in I$. Namely, if $g \in \mathrm{Ker}(\chi_i)$, for all $i \in I$. □

Lemma 16.1.3. *Any normal subgroup $N \lhd G$ is of the form $\mathrm{Ker}(\chi)$ for some character χ.*

Proof. Let $H = G/N$ and consider the composition

$$G \xrightarrow{\pi} G/N = H \xrightarrow{\rho_H^{reg}} \mathrm{GL}(\mathbb{C}[H]).$$

Let $\rho = \rho_H^{reg} \circ \pi$. Since the regular representation ρ_H^{reg} of H is injective, it follows that $\mathrm{Ker}(\rho) = \mathrm{Ker}(\pi) = N$. Therefore,

$$N = \mathrm{Ker}(\chi_\rho).$$

□

We summarize our discussion in the following theorem.

Theorem 16.1.4. *Let $\chi_1,\dots,\chi_h$, $h = h(G)$, be the irreducible characters of G. Let*

$$N_i = \mathrm{Ker}(\chi_i).$$

Any normal subgroup N of G is of the form

$$N = \cap_{i\in I}\mathrm{Ker}(\chi_i),$$

for a suitable subset $I \subseteq \{1,2,\dots,h\}$. And, conversely, any such intersection is a normal subgroup of G.

Remark 16.1.5. The whole point is, of course, that we have a practical easy method to find all the normal subgroups of a group G from the character table. Note, also, that the theorem implies that any proper maximal normal subgroup of G is of the form $\mathrm{Ker}(\chi_i)$ for some i (although, the converse is not true; $\mathrm{Ker}(\chi_i)$ is often not a maximal normal subgroup).

Example 16.1.6. We illustrate the theorem using the character table of A_4. Recall that it is given by the following table, where in the last column we indicated the kernel of the character.

We conclude that A_4 has only one non-trivial normal subgroup, which is K.

	1 1	(123) 4	(132) 4	$(12)(34)$ 3	Ker
$\mathbb{1}$	1	1	1	1	A_4
χ_1	1	ω	ω^2	1	K
χ_2	1	ω^2	ω	1	K
χ	3	0	0	-1	$\{1\}$

Table 9: Character table of A_4

16.2. **Recognizing the commutator subgroup.** Given a group G we have several normal subgroups canonically associated to it. For example, the commutator subgroup G' and the centre $Z(G)$. In light of Theorem 16.1.4, it makes sense to ask how to construct them from the character table. For the center, this is just the union of all conjugacy classes of size 1. For the commutator subgroup we have the following proposition.

Proposition 16.2.1. *We have*

$$G' = \bigcap_{\chi \text{ 1-dim. char.}} \mathrm{Ker}(\chi).$$

Proof. Suppose that $g \in G'$ and ρ is a 1-dimensional representation, then $\rho(G') = 1$ (and so, as we have used several times before, ρ factors through G^{ab}). Thus, $G' \subseteq \bigcap_{\chi \text{ 1-dim. char.}} \mathrm{Ker}(\chi)$.

Suppose now that $g \notin G'$ and denote $\bar{g}$ its image in G^{ab}. Then $\bar{g} \neq 0$ (the identity element of the abelian group G^{ab}). Write

$$G^{ab} \cong \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_a\mathbb{Z}.$$

Then $\bar{g} = (g_1, \ldots, g_a)$ and assume without loss of generality that $g_1 \neq 0$.

Let $\zeta = e^{2\pi i/n_1}$ and ρ the multiplicative character of $\mathbb{Z}/n_1\mathbb{Z}$ given by $\rho(a) = \zeta^a$. Then, $\rho \times \mathbb{1} \times \cdots \times \mathbb{1}$ is a multiplicative character of G^{ab} and hence, through $G \to G^{ab}$, also of G. We have

$$(\rho \times \mathbb{1} \times \cdots \times \mathbb{1})(g) = (\rho \times \mathbb{1} \times \cdots \times \mathbb{1})(\bar{g}) = \rho(g_1) = \zeta^{g_1} \neq 1.$$

Thus, $g \notin \bigcap_{\chi \text{ 1-dim. char.}} \mathrm{Ker}(\chi)$, and the proof is complete. □

17. More examples of representations

In this section we consider two more examples, more involved than those considered thus far.

17.1. **Character table of the Frobenius group F_{20}.** The **Frobenius group** F_{20} is the group

$$\mathbb{Z}/5\mathbb{Z} \rtimes (\mathbb{Z}/5\mathbb{Z})^\times.$$

Recall that $(\mathbb{Z}/5\mathbb{Z})^\times = \mathrm{Aut}(\mathbb{Z}/5\mathbb{Z})$ and the semi-direct product is taken relative to the identity map $(\mathbb{Z}/5\mathbb{Z})^\times \to \mathrm{Aut}(\mathbb{Z}/5\mathbb{Z})$. The group law is very simple,

$$(n_1, b_1)(n_2, b_2) = (n_1 + b_1 n_2, b_1 b_2), \qquad n_i \in N := \mathbb{Z}/5\mathbb{Z}, b_i \in B := (\mathbb{Z}/5\mathbb{Z})^\times.$$

The Frobenius group can be realized into other ways:

(1) As a group of matrices

$$\left\{ \left(\begin{smallmatrix} b & n \\ & 1 \end{smallmatrix}\right) : b \in \mathbb{Z}/5\mathbb{Z}^\times, n \in \mathbb{Z}/5\mathbb{Z} \right\},$$

with multiplication

$$\left(\begin{smallmatrix} b_1 & n_1 \\ & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} b_2 & n_2 \\ & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} b_1 b_2 & n_1 + b_1 n_2 \\ & 1 \end{smallmatrix}\right).$$

(2) As the subgroup of S_5 given by

$$\langle (12345), (2354) \rangle.$$

The isomorphism of F_{20} with the group of matrices is evident. For the realization as a subgroup of permutations we send

$$(12345)^n \mapsto (n,1), \qquad (2345) \mapsto (0,2).$$

Because

$$(2354)(12345)(2354)^{-1} = (12345)^2,$$

and $(0,2)(1,1)(0,2)^{-1} = (2,1)$, it follows (with some additional arguments) that we have an isomorphism $\langle (12345), (2354) \rangle \cong F_{20}$.

Next, we calculate the conjugacy classes of F_{20}. For elements of N, conjugation by N is trivial and so by conjugating by elements of B we get the full conjugacy classes (using that $F_{20} = NB$). We have the formula

$$(0,b)(n,1)(0,b^{-1}) = (bn,1).$$

We find two conjugacy classes:

$$a_1 = \{(0,1)\}, \quad a_2 = \{(i,1) : i = 1,2,3,4\}.$$

Likewise, conjugating elements of B by B is trivial and so we will get the full conjugacy classes of elements of B by conjugating them by elements of N. We have the relation

$$(n,1)(0,b)(-n,1) = ((1-b)n, b).$$

For $b = 2,3,4$, we get the conjugacy classes

$$c_2 = \{(i,2) : 0 \le i \le 4\}, \quad c_3 = \{(i,3) : 0 \le i \le 4\}, \quad c_4 = \{(i,4) : 0 \le i \le 4\}.$$

We see that we already accounted for all the elements of the group. Therefore, F_{20} has 5 conjugacy classes (of sizes $1,4,5,5,5$).

Note that $F_{20}/N \cong B \cong (\mathbb{Z}/5\mathbb{Z})^\times$, $(n,b) \mapsto b$. As F_{20} is not abelian, and N has no non-trivial subgroups, it follows that $N = F_{20}'$ and $F_{20}^{ab} \cong (\mathbb{Z}/5\mathbb{Z})^\times$, which is cyclic group of order 4 with generator 2.Thus, F_{20} has precisely 5 irreducible representations, 4 of which are 1-dimensional. Therefore, as the size of the group is the sum of the squares of the dimensions of the irreducible representations, the remaining irreducible representation is 4-dimensional. We can find its character χ_4 by using that the weighted sum of the rows of the character table is the regular representation.

	a_1 1 (0, 1)	a_2 4 (1, 1)	c_2 5 (0,2)	c_3 5 (0, 3)	c_4 5 (0, 4)
$\chi_0 = \mathbb{1}$	1	1	1	1	1
χ_1	1	1	i	$-i$	-1
χ_2	1	1	-1	-1	1
χ_3	1	1	$-i$	i	-1
χ_4	4	-1	0	0	0

Table 10: Character table of F_{20}

It is not hard to check that under the realization of F_{20} as a subgroup of S_5 in fact $\chi_4 = \chi^{std,0}|_{F_{20}}$.

17.2. **Monomial representations.** Consider a finite group G acting on a non-empty finite set S. Construct a vector space V with basis $\{e_s : s \in S\}$; we have $\dim(V) = \sharp S$. There is a natural representation

$$\rho\colon G \to \mathrm{GL}(V), \quad \rho(g)(e_s) = e_{g*s}.$$

Such representations are called **monomial**. In fact, we have already seen at least two instances of this construction:

- When $S = G$, and G acts by left multiplication, we get $V = \mathbb{C}[G]$ and $\rho = \rho^{reg}$.
- When $G = S_n$, and $S = \{1, 2, \ldots, n\}$, we get $V = \mathbb{C}^n$ and $\rho = \rho^{std}$.

As in these two examples, it is easy to check that

$$\chi_\rho(g) = I(g) = \sharp \text{ fixed points of } g \text{ in } S.$$

Applying CFF on the one hand, and the projection formula on the other hand, we get

$$\frac{1}{\sharp G}\sum_{g\in G}\chi_\rho(g) = \sharp \text{ orbits of } G \text{ in } S = \dim(V^G). \tag{19}$$

One way one may get such actions, is by choosing a subgroup $B < G$ and letting $S = G/B$, the set of left cosets of B in G (in fact, any set S on which G acts is a union of such examples). The representation is called the **coset representation**.

To make the situation even more specific, assume that

$$G = N \rtimes_\phi B.$$

Therefore, $G = NB, N \cap B = \{1\}$. Then,

$$G/B = \{nB : n \in N\}.$$

We check that $gnB = nB \Leftrightarrow g \in nBn^{-1}$. But,

$$nBn^{-1} = \{(n,1)(1,b)(n^{-1},1) : b \in B\} = \{(n\phi_b(n)^{-1}, b) : b \in B\}.$$

If $g = (n_1, b) \in nBn^{-1}$ it means that g necessarily equals to $(n\phi_b(n)^{-1}, b)$ for some n. We conclude that

$$\chi((n_1,b)) = I((n_1,b)) = \sharp\{n \in N : n_1 = n\phi_b(n)^{-1}\}.$$

Continuing with a general analysis will require making more assumptions on ϕ. Instead, let us take the case of $F_{20} = \mathbb{Z}/5\mathbb{Z} \rtimes_{id} (\mathbb{Z}/5\mathbb{Z})^\times$. Here, $n_1 = n\phi_b(n)^{-1}$ is written in additive notation and the condition is $n_1 = (1-b)n$. Now,

- if $b \neq 1$ there is a unique solution to the equation $n_1 = (1-b)n$.
- if $b = 1$ and $n_1 = 0$ there are 5 solutions to the equation $n_1 = (1-b)n$.
- if $b = 1$ and $n_1 \neq 0$ there are no solutions to the equation $n_1 = (1-b)n$.

We conclude that the character χ has the values $\chi(a_1) = 5, \chi(a_2) = 0, \chi(c_2) = \chi(c_3) = \chi(c_4) = 1$. Therefore,

$$\chi = \chi_4 + \chi_0,$$

and that tells us how the representation decomposes. Incidentally, note that the action of F_{20} on the 5 cosets of B gives us the inclusion $F_{20} \subset S_5$ we used before.

17.3. **A combinatorial application.** Let G be a finite group acting transitively on a finite non-empty set S. Let

$$G_0 = \{g \in G : g \text{ has no fixed point in } S\}.$$

G_0 is a subset of G, not a subgroup. It is not hard to prove (for example by using the Cauchy-Frobenius formula for the number of orbits for a finite group acting on a set) that if $\sharp S \geq 2$ then

$$\sharp G_0 \geq 1.$$

Theorem 17.3.1 (Cameron-Cohen)**.**

$$\sharp G_0 \geq \frac{\sharp G}{\sharp S}.$$

Proof. Let $I(g) = \chi(g)$ be the number of fixed points of g in S, where χ is the character of the monomial representation of G coming from S.

Compare the proof of the following lemma to the proof of Lemma 12.3.1. It is really the same.

Lemma 17.3.2. *We have*

$$\frac{1}{\sharp G}\sum_{g\in G}\chi^2(g) \geq 2.$$

Proof. Consider the action of G on the set $S \times S$, $g(a,b) = (g(a), g(b))$. The class function χ^2 is the character of this representation and the dimension of the space of invariant vectors is $\frac{1}{\sharp G}\sum_{g\in G}\chi^2(g)$, which is equal to the number of orbits of G in $S \times S$ by Equation (19). To prove the lemma we only need to show that there is more than 1 orbit. And, indeed, since we assumed that G acts transitively on S, one orbit is the diagonal $\{(s,s) : s \in S\}$; since $\|S\| \geq 2$, there must be at least one more orbit. □

Let $n = \sharp S$. Note that for $g \notin G_0$ we have $1 \leq \chi(g) \leq n$ and therefore

$$\frac{1}{\sharp G}\sum_{g\in G-G_0}(\chi(g)-1)(\chi(g)-n) \leq 0.$$

Therefore,

$$\frac{1}{\sharp G}\sum_{g\in G}(\chi(g)-1)(\chi(g)-n) \leq \frac{1}{\sharp G}\sum_{g\in G_0}(\chi(g)-1)(\chi(g)-n) = n\cdot\frac{\sharp G_0}{\sharp G}.$$

On the other hand,

$$\frac{1}{\sharp G}\sum_{g\in G}(\chi(g)-1)(\chi(g)-n) = \frac{1}{\sharp G}\sum_{g\in G}\chi^2(g) - (n+1)\frac{1}{\sharp G}\sum_{g\in G}\chi(g) + \frac{1}{\sharp G}\sum_{g\in G}n$$
$$\geq 2-(n+1)+n = 1.$$

Combining the two inequalities, the theorem follows. □

The proof of the following theorem in number theory is beyond the scope of this course. It uses the Cameron-Cohen Theorem.

Theorem 17.3.3. (J.-P. Serre) *Let $f(x) \in \mathbb{Z}[x]$ be an irreducible polynomial of degree n. The density of prime numbers p (in the set of all primes) such that f has no root modulo p is at least $1/n$.*

Example 17.3.4. If we take the most simple non-trivial situation $f(x) = x^2+1$, the theorem states that for at least $1/2$ the primes f has no zero modulo p.

On the other hand, f has a zero modulo p if and only if -1 is a square modulo p. As -1 has order 2 modulo p (if $p > 2$), this happens if and only if there are elements of order 4 in $\mathbb{Z}/p\mathbb{Z}^\times$. Using that $\mathbb{Z}/p\mathbb{Z}^\times$ is a cyclic group of order $p-1$ we see that this is the case if and only if $p \equiv 1 \pmod 4$. Thus, we conclude that the density of primes of the form $4k+3$ is at least $\frac{1}{2}$ (in fact, it is known to be precisely $1/2$).

18. The theorems of Burnside and Blichfeldt

In this section we prove three important results. The first is that the dimension of an irreducible representation divides the order of the group. We then prove Burnside's theorem that groups of order $p^a q^b$ are solvable, using group representations. Finally, we prove a theorem of Blichfeldt that describes every representation of a supersolvable group G as an induction of a multiplicative character (i.e., of a 1-dimensional representation) from a subgroup, thus providing an algorithm to construct every irreducible representation. This theory can be extended further; we mention below a theorem of Brauer in this context that applies for a wider class of groups, but we will not prove it here.

18.1. **Dimensions of irreducible representations.**

Theorem 18.1.1. *Let ρ be an irreducible representation of a finite group G then*

$$\dim(\rho) | \sharp\, G.$$

Proof. We first note that if χ is a character of G then as $\chi(g)$ is a sum of roots of unity, which are algebraic integers, for all $g \in G$ also $\chi(g)$ is an algebraic integer. Let $g \in G$ and let C be the conjugacy class of g in G. Let (ρ, V) be an irreducible representation of G with character χ. Consider the operator

$$\rho(C) := \sum_{x \in C} \rho(x) \colon V \to V.$$

Note $\rho(C)$ is equivariant for the action of G. That is, $\rho(h) \circ \rho(C) \circ \rho(h^{-1}) = \rho(C)$, which is easy to verify using that $hCh^{-1} = C$. Thus, by Schur's lemma, $\rho(C)$ is a scalar operator and so it is equal to $\lambda \cdot Id$, where

$$\lambda = \frac{1}{\dim(V)} \mathrm{Tr}(\rho(C)) = \frac{\chi(g) \sharp C}{\dim(V)} = \frac{\chi(g) \sharp C}{\chi(1)}$$

(using that χ is constant along C). Note that

$$\frac{\chi(g) \sharp C}{\chi(1)}$$

is an algebraic *number* (not known yet to be an algebraic integer).

To show that $\frac{\chi(g) \sharp C}{\chi(1)}$ is an algebraic *integer*, we realize V as a sub-representation of the regular representation $\mathbb{C}[G]$. This is possible because the regular representation contains every irreducible representation. Then the action of $g \in G$ on an element $u \in V$ is just gu, multiplication in the group ring $\mathbb{C}[G]$.

Lemma 18.1.2. *Let $r := \sum_{g \in G} a_g [g]$ be an element of $\mathbb{C}[G]$ such that a_g is an integer for all g. Suppose that $(\sum_{g \in G} a_g [g]) u = \lambda u$, for some algebraic number λ and non-zero $u \in \mathbb{C}[G]$. Then λ is an algebraic integer.*

Proof of Lemma. Left-multiplication by r defines a $\mathbb{C}$-linear transformation

$$\mathbb{C}[G] \to \mathbb{C}[G], \quad u \mapsto ru.$$

Let us number the elements in G by $\{1 = g_1, \ldots, g_n\}$ and so $r = \sum_{i=1}^n a_i [g_i]$. Relative to the $\mathbb{C}$-basis $\{[g_i] : 1 \leq i \leq n\}$ for $\mathbb{C}[G]$, the left-multiplication-by-r map is represented by a matrix with integral coefficients because it sends the j-th basis element $[g_j]$ to $\sum_i a_i \cdot [g_i g_j]$. Which means that the j column of the matrix is comprised the elements $\{a_i\}$ in some permuted order.

Since the characteristic polynomial of a matrix with integer coefficients is a monic polynomial with integer coefficients, any eigenvalue λ of the matrix, being a root of the characteristic polynomial, is an algebraic integer. $\square$

Returning to the proof of the theorem, note that $\rho(C)$ is an element as in the lemma with $a_g = 1$ if $g \in C$, and $a_g = 0$ otherwise. It then follows that

$$\frac{\chi(g)\sharp C}{\chi(1)} \tag{20}$$

is an algebraic *integer* (any non-zero element of the representation V is a corresponding eigenvector).

Now, let $h_1, \dots, h_r$ be representatives for the conjugacy classes $C_1, \dots, C_r$ of G. Then

$$\sum_{i=1}^{r} \frac{\chi(h_i)\sharp C_i}{\chi(1)} \cdot \bar{\chi}(h_i)$$

is an algebraic integer. But this algebraic integer is equal to

$$\frac{\sharp G}{\chi(1)} \langle \chi, \chi \rangle = \frac{\sharp G}{\chi(1)} \in \mathbb{Q}.$$

Therefore, $\frac{\sharp G}{\chi(1)}$ is an integer, meaning $\chi(1) = \dim(V) | \sharp G$. □

18.2. **Burnside's Theorem.**

Theorem 18.2.1. (Burnside) *Let $p \neq q$ be primes and a, b be non-negative integers. Any group of order $p^a q^b$ is solvable.*

We first prove the following theorem, which is of independent interest.

Theorem 18.2.2. *Let G be a finite group and p a prime number. Suppose that G has a conjugacy class of order p^r for some positive integer r. Then G is not simple.*

Proof. This is one of those proofs that are correct, but it's really hard to figure out what is going on. It just works.

First note that the group G cannot be abelian and if the conjugacy class C of order p^r is the conjugacy class of g then g cannot be the identity element.

Let $\chi_1 = \mathbb{1}, \dots, \chi_h$ be the irreducible characters of G. We use the orthogonality of the columns in the character table of G for the distinct columns of 1_G and g, to conclude that

$$\sum_{i=1}^{h} \chi_i(g)\bar{\chi}_i(1) = 1 + \sum_{i=2}^{h} \chi_i(g)\chi_i(1) = 0.$$

This implies that $\sum_{i=2}^{h} \chi_i(g) \cdot \frac{\chi_i(1)}{p} = -1/p$ is not an algebraic integer and so, for some $i \geq 2$ for which $\chi_i(g) \neq 0$, $\frac{\chi_i(1)}{p}$ is not an algebraic integer. Namely, for some i we have

$$\chi_i(g) \neq 0, \qquad p \nmid \chi_i(1).$$

Now, for suitable integers a, b we have $a \cdot \sharp C + b \cdot \chi_i(1) = 1$. Multiply this by $\chi_i(g)/\chi_i(1)$ to find

$$a \cdot \frac{\chi_i(g)\sharp C}{\chi_i(1)} + b\chi_i(g) = \frac{\chi_i(g)}{\chi_i(1)}.$$

Using this equation and (20), we find that $\frac{\chi_i(g)}{\chi_i(1)}$ is an algebraic integer. We claim that

$$\left| \frac{\chi_i(g)}{\chi_i(1)} \right| = 1.$$

Indeed, denote $\gamma = \frac{\chi_i(g)}{\chi_i(1)} \neq 0$. As $\chi_i(g)$ is a sum of d roots of unity, say $\chi_i(g) = \zeta_1 + \cdots + \zeta_d$, where $d = \chi_i(1)$, we have $|\gamma| \leq 1$. Suppose that $|\gamma| < 1$. The minimal polynomial of γ has roots $(\sigma(\zeta_1) + \cdots + \sigma(\zeta_d))/d$, where $\sigma \in \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$; each such root has absolute value at most 1 and one of them, e.g. γ, has

absolute value less than 1. It follows that the constant coefficient of the minimal polynomial, which is the product of the roots, has absolute less than 1, yet it is a non-zero integer because γ is a non-zero algebraic integer. This is a contradiction.

We know $\gamma = (\zeta_1 + \cdots + \zeta_d)/d$ has absolute value 1. As each ζ_i has absolute value 1, this is possible if and only if all the ζ_i are pointing in the same direction and therefore are equal; that is to say, $\rho_i(g)$ is a scalar. As such, $\rho_i(g)$ commutes with all $\rho_i(h), h \in G$.

Now, if ρ_i is *not* a faithful representation, its kernel is a non-trivial normal subgroup of G which is not G (because $\chi_i \neq \chi_1$), showing G is not simple.

If ρ_i is faithful, we may conclude that g commutes with all $h \in G$ and so $g \in Z(G)$. This shows $Z(G) \neq \{1\}$ and, as G is not abelian, $Z(G)$ is a non-trivial proper normal subgroup of G and again we find that G is not simple. □

Proof. (Of Burnside's Theorem) Arguing by induction, it is enough to prove that such a group is not simple. Moreover, given basic results about p-groups, we may assume that a and b are positive and G is not abelian.

Let Q be a q-Sylow subgroup. Note that $Z(Q) \neq \{1\}$ and take $1 \neq g \in Z(Q)$. Then $\mathbf{Cent}(g) \supset Q$ and thus $\mathbf{Conj}(g)$ is of cardinality p^r. If $r = 0$, $g \in Z(G)$ and so $Z(G)$ is a non-trivial normal subgroup of G. If $r \geq 1$, G is not simple by Theorem 18.2.2. □

18.3. **Blichfeldt's theorem.** Blichfeldt's theorem is a rather striking result that asserts that for supersolvable groups, in particular, for p-groups, every irreducible representation is induced from a 1-dimensional representation of a subgroup.

Let G be a finite group. We say that G is **supersolvable** if there is a sequence of subgroups

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_N = \{1\}, \tag{21}$$

of subgroups G_i, such that each G_i is a normal subgroup of G (not just of G_{i-1}!) and such that G_{i-1}/G_i is cyclic for $i = 1, \ldots, N$. For example, every abelian group and every p-group is supersolvable.

It is an exercise to show that we may, equivalently, require that each G_{i-1}/G_i is cyclic of prime order. Namely, that given a series such as in (21), one can refine the series so that the quotients are cyclic of prime order and still every subgroup is normal in G.

It is a standard argument to show that just like for solvable groups, subgroups and quotient groups of supersolvable groups are supersolvable. However, unlike in the case of solvable groups, it is not true that an extension of a supersolvable group by a supersolvable group is supersolvable. For example, the group S_4 sits in an exact sequence $1 \to K \to S_4 \to S_3 \to 1$, where K (the Kline group) and S_3 are supersolvable, but S_4 is not supersolvable: S_4 doesn't have any normal subgroup of order 2 or 3. We see that

$$\text{supersolvable} \underset{\not\Longleftarrow}{\Longrightarrow} \text{solvable.}$$

However, a direct product of supersolvable groups is supersolvable.

Theorem 18.3.1 (Blichfeldt)**.** *Let G be a supersolvable group and let ρ be an irreducible representation of G then*

$$\rho \cong \mathrm{Ind}_J^G \psi,$$

where J is a subgroup of G and ψ a 1-dimensional representation of J.

Proof. The proof is a mix of induction on the order of G and also of reduction to the faithful case. The base of the induction is the case where G is cyclic of prime order in which case any irreducible representation is one-dimensional and so we may take $J = G$.

Now for the general case: we first assume that G acts faithfully on V. Namely, that

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

is an injective map.

First note that if G is abelian there is nothing to prove; ρ must be one dimensional and we take $J = G$.

Lemma 18.3.2. *Let G be a non-abelian supersolvable group. Then G has an abelian normal subgroup N such that $N \not\subseteq Z(G)$.*

Proof. (Lemma) Let G_i be the minimal subgroup in (21) such that $G_i \not\subseteq Z(G)$. Then $G_{i+1} \subseteq Z(G)$ and G_i/G_{i+1} is cyclic. Then G_i is a normal subgroup and is abelian: every element of G_i can be written in the form $x^n y$ for some fixed $x \in G_i$ and $y \in G_{i+1}$. As elements of G_{i+1} commute with elements of G_i, in particular with x, we find for two elements $x^{n_1}y_1, x^{n_2}y_2$ that

$$x^{n_1}y_1x^{n_2}y_2 = x^{n_1+n_2}y_1y_2 = x^{n_1+n_2}y_2y_1 = x^{n_2}y_2x^{n_1}y_1.$$

□

Fix this subgroup N and consider V as a representation of N. As N is abelian, V decomposes as a direct sum over the character group of N:

$$V = \oplus_{\psi \in N^*} V_\psi,$$

where $V_\psi = \{v \in V : \rho(n)v = \psi(n)v, \forall n \in N\}$. Pick a ψ such that $V_\psi \neq \{0\}$. For $g \in G$ let

$$\psi^g : N \to \mathbb{C}^\times, \qquad \psi^g(n) = \psi(g^{-1}ng).$$

These are characters of N. Let $S = \{\psi^g : g \in G\} \subseteq N^*$. Let $H = \{g \in G : \psi^g = \psi\}$. Then,

$$\sharp S = [G : H].$$

Now, it is easy to check that

$$\rho(g)(V_\chi) = V_{\chi^g}.$$

Thus, in fact,

$$V = \oplus_{\chi \in S} V_\chi.$$

Note that for every $\chi \in S$ we have an isomorphism of vector spaces $V_\chi \cong V_\psi$ (but not as representations) and that give

$$\dim(V) = \sharp S \cdot \dim(V_\psi).$$

Now consider the map of representations, given on pure tensors by $g \otimes v \mapsto \rho(g)(v)$,

$$\mathrm{Ind}_H^G V_\psi = \mathbb{C}[G] \otimes_{\mathbb{C}[H]} V_\psi \to V. \tag{22}$$

As this is not the zero map and V is irreducible, this map is surjective. The dimension of the l.h.s is $[G : H] \cdot \dim(V_\psi) = \sharp S \cdot \dim(V_\psi) = \dim(V)$ and therefore the map is an isomorphism.

In fact, we claim that S has more than one element (and consequently H is a proper subgroup of G). Indeed, if S has only one element then $\psi^g = \psi$ for all $g \in G$ and $V = V_\psi$. That is, for all $n \in N$ and $g \in G$ we have

$$\rho(g^{-1}ngn^{-1}) = \psi(g^{-1}ngn^{-1}) = 1.$$

As ρ is faithful, this implies that $g^{-1}ngn^{-1} = 1$ for all $n \in N, g \in G$. But that implies $N \subseteq Z(G)$. Contradiction!

Thus, we have proven that $V \cong \mathrm{Ind}_H^G V_\psi$ for some proper subgroup H of G. Note that V_ψ is an irreducible representation of H, else the isomorphism in (22) would imply that V is reducible too. As H is a supersolvable subgroup of smaller order, we may apply the induction hypothesis to conclude that

$$V_\psi \cong \mathrm{Ind}_J^H U,$$

where U is a 1-dimensional representation of a subgroup J of H. Then

$$V \cong \mathrm{Ind}_H^G \mathrm{Ind}_J^H U \cong \mathrm{Ind}_J^G U,$$

and the proof is complete for the case where G acts faithfully.

Suppose that G does not act faithfully and let $G_0 = \mathrm{Ker}(\rho)$. Using the first isomorphism theorem, there is an injective homomorphism $\rho_0 \colon G/G_0 \to GL(V)$ such that ρ is the composition $G \to G/G_0 \xrightarrow{\rho_0} \mathrm{GL}(V)$. As G/G_0 is a supersolvable group of smaller order, we can apply induction and conclude that

$$\rho_0 \cong \mathrm{Ind}_{J'}^{G/G_0} U,$$

for some one dimensional representation U of a subgroup J' of G/G_0 (where we view both sides also as representations of G). Let J be the preimage of J' under the homomorphism $G \to G/G_0$. Then, viewing U as a 1-dimensional representation of J, one finds, for example by calculating characters,

$$\rho \cong \mathrm{Ind}_J^G U.$$

□

Example 18.3.3. Let consider again the Frobenius group F_{20} of §17.1:

$$\mathbb{Z}/5\mathbb{Z} \rtimes (\mathbb{Z}/5\mathbb{Z})^\times.$$

Let $N = \mathbb{Z}/5\mathbb{Z} \lhd F_{20}$. One checks that any non-trivial normal subgroup of F_{20} contains N and that F_{20} has no centre. Any irreducible representation of F_{20} of dimension greater than 1 is faithful, else it comes from a representation of a quotient group of F_{20}, but all these quotient groups are abelian. Thus, N is the group to take in the proof of Blichfeldt's theorem. One deduces that every irreducible representation of F_{20} is induced from a multiplicative character of N. Fixing a fifth root of unity ζ, the multiplicative characters of N are

$$\chi_a(b) = \zeta^{ab}, \quad a = 0, 1, 2, 3, 4.$$

We claim that $\mathrm{Ind}_N^{F_{20}} \chi_a$ is the same irreducible 4-dimensional representation of F_{20} if $a \neq 0$. We use Frobenius reciprocity, with the simplification afforded by the normality of N (Corollary 13.1.2)

$$\mathrm{Ind}_N^G \chi_a(g) = \begin{cases} 0 & g \notin N \\ \frac{1}{|N|} \sum_{b \in G} \chi_a(b^{-1} g b) & g \in N. \end{cases}$$

Note that for $g \in N$, each $b^{-1}gb$ appears the same number of times, $\sharp\mathrm{Cent}(g) = \sharp N$, when b ranges of G. Also, if $g \neq 1$, it is not hard to check that the conjugacy class of g is $N - \{0\}$. Thus, for $g \neq 0$,

$$\frac{1}{|N|} \sum_{b \in G} \chi_a(b^{-1} g b) = \chi_a(1) + \chi_a(2) + \chi_a(3) + \chi_a(4).$$

This sum is equal to -1 if $a \neq 0$ and is equal to 4 for $a = 0$. Thus, all the induced characters for $a \neq 0$ are the same, implying that the induced representations are isomorphic. And,

$$\langle \mathrm{Ind}_N^G \chi_a, \mathrm{Ind}_N^G \chi_a \rangle = \langle \chi_a, \mathrm{Res}_N^G \mathrm{Ind}_N^G \chi_a \rangle = \frac{1}{|N|}(4 - \sum_{b=1}^{4} \zeta^{ab}) = 1.$$

Namely, each such induced representation is irreducible. In contrast, when $a = 0$, the same calculation gives $\|\mathrm{Ind}_N^G \mathbb{1}\|^2 = 4$, showing that the induced representation is reducible.

Example 18.3.4. This is a generalization of the previous example, but a bit rephrased. Let $p > 2$ be a prime number, and let

$$F = F_{p(p-1)} = \mathbb{Z}/p\mathbb{Z} \rtimes (\mathbb{Z}/p\mathbb{Z})^\times.$$

We let $N = \mathbb{Z}/p\mathbb{Z}$, viewed as a subgroup of F via $n \mapsto (n, 1)$ and we also view $(\mathbb{Z}/p\mathbb{Z})^\times$ as a subgroup via $b \mapsto (0, b)$. The group law in F is

$$(a_1, b_2)(a_2, b_2) = (a_1 + b_2 a_2, b_1 b_2).$$

The series

$$F \supset N \supset \{1\},$$

has cyclic quotients, namely $\mathbb{Z}/p\mathbb{Z}^\times$ and $\mathbb{Z}/p\mathbb{Z}$ (the multiplicative group of a finite field is cyclic), and hence F is supersolvable.

The following properties of F are not hard to check and are left as exercises:

- $Z(F) = \{1\}$, $F' = N$.
- Any non-trivial normal subgroup K of F contains N.

We conclude that F has $p-1$ one-dimensional representations, all obtained as a composition

$$F \to \mathbb{Z}/p\mathbb{Z}^\times \to \mathbb{C}^\times.$$

Also, if (ρ, V) is an irreducible representation of F of dimension greater than 1 then ρ is faithful, else ρ is obtained as a composition $F \twoheadrightarrow F/K \xrightarrow{\rho_0} \mathrm{GL}(V)$, where ρ_0 is irreducible. But $K \supseteq N$ and so F/K is abelian, implying that ρ_0 (and hence ρ) is 1-dimensional. Thus, following the proof of Blichfeldt's theorem, the N is our N, and we are in the faithful case. The trivial character $\mathbb{1}$ of N always induces a reducible representation of F – it is a monomial representation on the set of cosets of N, of dimension $p-1 > 1$, that has the trivial representation as a subrepresentation. Thus, in the proof we have some $\psi : N \to \mathbb{C}^\times$ such that $V_\psi \neq \{0\}$, $\psi(1) = \zeta$, where $\zeta \neq 1, \zeta^p = 1$.

The action of N on ψ by twisting is trivial and the action of $b \in \mathbb{Z}/p\mathbb{Z}^\times$ can be check to be

$$\psi \mapsto \psi^b, \quad \psi^b : N \to \mathbb{C}^\times, \quad \psi^b(n) = \psi(bn) = \zeta^{bn}.$$

In particular, we find that H in the proof is just N and every non-trivial character of N appears in the decomposition of V:

$$V = \oplus_{\chi \in N^*, \chi \neq \mathbb{1}} V_\chi \cong \mathrm{Ind}_N^F \psi,$$

for any choice of a non-trivial character ψ of F. In particular, F has a unique irreducible representation of dimension greater than 1, it is induced from any non-trivial character of N and has dimension $p-1$. We check that

$$p(p-1) = (p-1) \cdot 1^2 + (p-1)^2.$$

Remark 18.3.5. There is an important theorem, in the same circle of ideas, that should be mentioned. To state it, we will need some additional terminology. Consider the subring

$$\mathrm{ch}(\mathbb{C}[G])$$

of $\mathbb{C}[G]$ consisting of all $\mathbb{Z}$-linear combinations of the irreducible characters $\{\chi_i : i = 1, \dots, h(G)\}$. An element f of $Class(G)$ lies in $\mathrm{ch}(\mathbb{C}[G])$ if and only if $\langle f, \chi_i \rangle \in \mathbb{Z}, i = 1, \dots h(G)$. The elements of $\mathrm{ch}(\mathbb{C}[G])$ are called **virtual characters**. They are class functions f that are good candidates to be characters of representations, especially if $f(1) \geq 0$, but this condition doesn't suffice.

Let p be a prime. A group H is called p-**elementary** if H is a direct product of a p-group with a cyclic subgroup of order prime to p. It is called **elementary** if it is p-elementary for some prime p. Given a group G we can consider the family of all its elementary subgroups.

Theorem 18.3.6. **(Brauer's Induction Theorem)** *Every virtual character of G is a $\mathbb{Z}$-linear combination of characters induced from 1-dimensional characters of elementary subgroups of G. That is, any element $f \in \mathrm{ch}(\mathbb{C}[G])$ is of the form*

$$f = \sum_{H_i} a_i \cdot \mathrm{Ind}_{H_i}^G \psi_i,$$

for some integers a_i, elementary subgroups H_i of G and 1-dimensional characters $\psi_i : H_i \to \mathbb{C}^\times$.

Brauer's motivation was to prove this way that L-functions constructed in number theory have meromorphic continuation to the complex plane. This factorization formula expresses an L function associated to a Galois extension of $\mathbb{Q}$ with Galois group G as a product of Dirichlet L-functions to powers a_i. For Dirichlet L-function one knows analytic continuation and so, if one knows all the a_i are positive, one would even get holomorphic continuation, which is still an open problem known as **Artin's Conjecture**.

19. Further operations on representations

We have seen already the construction of the dual representation and the tensor product of representations. We provide a few more constructions of representations of linear algebra nature.

19.1. **Symmetric and alternating products.** The importance of symmetric products and alternating products goes much beyond representations of groups and so it is well worth to spend some time on these notions.

19.1.1. *Graded rings.* Let R be a ring. R is called a **graded ring** if R decomposes as a direct sum of abelian groups

$$R = \oplus_{n=0}^{\infty} R_n,$$

such that for all m, n,

$$R_m R_n \subseteq R_{n+m}.$$

Note that R_0 is a subring of R. A two-sided ideal $I \lhd R$ is called **graded** if

$$I = \oplus_{n=0}^{\infty} I_n, \quad I_n \subseteq R_n.$$

In this case, the quotient ring R/I is also graded as

$$R/I \cong \oplus_{n=0}^{\infty} R_n/I_n.$$

It is not hard to prove that if r_i are elements R such that each $r_i \in R_{n(i)}$ then the two sided ideal I generated by them $\langle \{r_i\} \rangle$ is a graded ideal. Sometimes we will use the notation $R = \oplus_{n=0}^{\infty} R^n$ instead of $R = \oplus_{n=0}^{\infty} R_n$.

19.1.2. *Tensor and symmetric algebras.* Let R be a commutative ring and let V be an R-module. We define

$$T^\bullet(V) = \oplus_{n=0}^{\infty} T^n(V),$$

where

$$T^0(V) = R, \quad T^1(V) = V, \quad T^n(V) = V \otimes_R V \otimes_R \cdots \otimes_R V \ \ (n - \text{times}).$$

$T^\bullet(V)$ is a graded R-algebra: for every m and n there is a natural map [24]

$$T^m(V) \times T^n(V) \to T^{m+n}(V), \quad (v_1 \otimes \cdots \otimes v_m, w_1 \otimes \cdots \otimes w_n) \mapsto v_1 \otimes \cdots \otimes v_m \otimes w_1 \otimes \cdots \otimes w_n.$$

[24] To be honest, to show that this map is well-defined requires an argument. For that we assume that $T^n(V)$ "solves" the problem of R-multiadditive, or even R-multilinear, maps $V \times \cdots \times V \to W$ for any W, in the same way that $V \otimes V$ "solves" the problem of R-biadditive, or bilinear, maps. See also § 19.2. For a fixed $(w_1, \cdots, w_n)$, the map

$$V^m \to T^{m+n}, \quad (v_1, \ldots, v_m) \mapsto v_1 \otimes \cdots \otimes v_m \otimes w_1 \otimes \cdots \otimes w_n,$$

is an R-multilinear map. Thus, we get a well-defined homomorphism of R-modules

$$\varphi_{(w_1,\ldots,w_n)} \colon T^m \to T^{m+n}, \quad v_1 \otimes \cdots \otimes v_m \mapsto v_1 \otimes \cdots \otimes v_m \otimes w_1 \otimes \cdots \otimes w_n.$$

This provides an R-multilinear map

$$V^n \to \mathrm{Hom}(T^m, T^{n+m}), \quad (w_1, \ldots, w_n) \mapsto \varphi_{(w_1,\ldots,w_n)},$$

and one concludes a homomorphism

$$T^n(V) \to \mathrm{Hom}_R(T^m, T^{n+m}),$$

given on pure tensors by

$$w_1 \otimes \cdots \otimes w_n \mapsto \varphi_{(w_1,\ldots,w_n)}.$$

We finally get a well-defined map

$$T^m(V) \times T^n(V) \to T^{m+n}(V),$$

This extends to define a product law on $T^\bullet(V)$, the **tensor algebra** of V.

Assume that V is a free R-module of rank d with a basis $e_1,\dots,e_d$, then $T^n(V)$ is a free R-module of rank d^n with the basis

$$\{e_{i_1}\otimes\cdots\otimes e_{i_n} : 1\le i_j\le d\}.$$

We let I be the two-sided ideal generated by all the tensors $\{x\otimes y - y\otimes x : x,y\in V\}$ where $1\le i,j\le d$. Then I is a graded ideal

$$I=\oplus_{n=0}^\infty I_n,$$

where, in fact $I_0=I_1=0$, I_2 is the R-span of the elements $\{x\otimes y - y\otimes x : x,y\in V\}$ and I_n is the R-span of the elements

$$\{x_1\otimes x_2\otimes\cdots\otimes x_n - x_{\sigma(1)}\otimes x_{\sigma(2)}\otimes\cdots\otimes x_{\sigma(n)} : x_i\in V,\sigma\in S_n\}.$$

We define the **symmetric algebra**

$$\boxed{\mathrm{Sym}^\bullet(V)=T^\bullet(V)/I=\oplus_{n=0}^\infty T^n/I_n,}$$

and denote its graded pieces by

$$\mathrm{Sym}^n(V):=T^n(V)/I_n.$$

We shall denote the image of the tensor $v_1\otimes\cdots\otimes v_n$ of $V^{\otimes n}$ in $\mathrm{Sym}^n(V)$ by $v_1\cdots v_n$ (remembering that we are allowed now to switch the order of the v_i as we please).

Suppose that $R=\mathbb{C}$. If G acts on the finite dimensional vector space V linearly, then it acts linearly on each $T^n(V)$, and so on $T^\bullet(V)$, by

$$g\cdot v_1\otimes\cdots\otimes v_n = gv_1\otimes gv_2\otimes\cdots\otimes gv_n.$$

The ideal I is a G-representation as well and so $\mathrm{Sym}^\bullet(V)$ is a graded G-representation and each graded piece is a finite dimensional representation of V. It is not hard to see that

$$\{e_{i_1}\otimes\cdots\otimes e_{i_n} : 1\le i_1\le\cdots\le i_n\le d\}$$

is a basis of $\mathrm{Sym}^n(V)$ and therefore

$$\dim(\mathrm{Sym}^n(V))=\binom{n+d-1}{n}.$$

Remark 19.1.1. Choose a basis $x_1,\dots,x_d$ for the $\mathbb{C}$-vector space V. It is not hard to see then that a basis for $V^{\otimes n}$ is

$$\{x_{i_1}x_{i_2}\cdots x_{i_n} : i_j\in\{1,\dots,d\}\}.$$

(It has cardinality d^n.) Thus, we may view $T^\bullet(V)$ as the ring of polynomials in the *non-commuting* variables $x_1,\dots,x_d$. To bring out this interpretation we have on purpose written $x_{i_1}x_{i_2}\cdots x_{i_n}$ instead of $x_{i_1}\otimes x_{i_2}\otimes\cdots\otimes x_{i_n}$. From this perspective, $\mathrm{Sym}^\bullet(V)$ can be interpreted as the ring of (usual) complex polynomials in the variables $x_1,\dots,x_n$ and a basis for $\mathrm{Sym}^n(V)$ is the given by the monomials

$$\{x_1^{i_1}\cdots x_d^{i_d} : i_j\ge 0, i_1+\cdots+i_d=n\},$$

and has cardinality $\binom{n+d-1}{n}$.

---

given on pure tensors by

$$(v_1\otimes\cdots\otimes v_m, w_1\otimes\cdots\otimes w_n)\mapsto\varphi_{(w_1,\dots,w_n)}(v_1\otimes\cdots\otimes v_m)=v_1\otimes\cdots\otimes v_m\otimes w_1\otimes\cdots\otimes w_n,$$

and the fact that it is an R-bilinear map.

19.1.3. *Example.* Let us deviate now from our usual conventions and allow G to be the infinite group $\mathrm{GL}_2(\mathbb{C})$. Let us take $V = \mathbb{C}^2$ the standard representation of $\mathrm{GL}_2(\mathbb{C})$, but let us think about it as linear forms $\alpha x + \beta y$, where $\alpha, \beta \in \mathbb{C}$. A matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ acts by sending x $(= {}^t(1,0))$ to $ax + cy$ $(= {}^t(a,b))$ and y to $bx + dy$. That is, the action amounts to a linear change of variables.

In this interpretation, an element of $\mathrm{Sym}^n(V)$ is a homogeneous polynomial $f(x,y)$ of degree n and the action is

$$f(x,y) \mapsto f(ax + cy, bx + dy).$$

We get therefore a series of representations of $\mathrm{GL}_2(\mathbb{C})$, $\{\mathrm{Sym}^n(V) : n = 0,1,2,\dots\}$. There is also the series of one dimensional representations $\{\det^n : n \in \mathbb{Z}\}$. Combined, we find that representations

$$\det{}^a \otimes \mathrm{Sym}^b(V), \qquad a \in \mathbb{Z}, b \in \mathbb{N}.$$

It turns out that all these representations are irreducible and non-isomorphic. Furthermore, every *algebraic* representation of $\mathrm{GL}_2(\mathbb{C})$, i.e., every homomorphism of groups $\mathrm{GL}_2(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$ given by rational functions $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto (f_{ij}(a,b,c,d))$, where the f_{ij} are of the form of a polynomial in a,b,c,d divided by some power of the determinant, is one of these representations. This gives a complete classification of the representations of $\mathrm{GL}_2(\mathbb{C})$. A similar theory exists for any algebraic group, or even any Lie group, in place of $\mathrm{GL}_2(\mathbb{C})$.

19.1.4. *The character of* Sym^2. Suppose that G acts on V with character χ. The character of the representation $T^2(V)$ is therefore χ^2. To calculate the character of $\mathrm{Sym}^2(V)$ take an element $g \in G$ and a basis $\{e_1, \dots, e_d\}$ for V on which it acts diagonally, say by $\mathbf{diag}(\alpha_1, \dots, \alpha_d)$. It acts then on $e_i e_j$ by $\alpha_i \alpha_j$ and so we find that the trace is $\sum_{i \le j} \alpha_i \alpha_j$. We arrive at the following formula

$$\chi_{\mathrm{Sym}^2(V)}(g) = \frac{1}{2}(\chi_V(g)^2 + \chi_V(g^2)). \tag{23}$$

Let us look at an example. Recall the character table of S_4.

	1	(12)	(123)	(1234)	$(12)(34)$
	1	6	8	6	3
$\mathbb{1}$	1	1	1	1	1
sgn	1	-1	1	-1	1
$\chi_3 \circ f$	2	0	-1	0	2
$\chi^{std,0}$	3	1	0	-1	-1
$\chi^{std,0} \cdot \mathrm{sgn}$	3	-1	0	1	-1

Let $\chi = \chi_3 \circ f$, the character of the irreducible-two dimensional representation. Then $\mathrm{Sym}^2(\chi)$ is the character of a 3-dimensional representation and is calculated using the table below. Using characters we find that $\mathrm{Sym}^2(\chi)$ is reducible and in fact

$$\mathrm{Sym}^2(\chi) = \chi_1 + \mathbb{1}.$$

	1	(12)	(123)	(1234)	$(12)(34)$
	1	6	8	6	3
χ	2	0	-1	0	2
χ^2	4	0	1	0	4
$\chi(g^2)$	2	2	-1	2	2
$\mathrm{Sym}^2(\chi)$	3	1	0	1	3

19.1.5. *The exterior algebra.* The exterior algebra is likewise constructed as a quotient of the tensor algebra by a graded ideal $J = \oplus_{n=0}^{\infty} J_n$. But now we want tensors to anti-commute. Thus, we want that under the map

$$T^n \to \bigwedge^n V := V^{\otimes n}/J_n,$$

that the image of $v_1 \otimes \cdots \otimes v_n$ is equal to the image of $\mathrm{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}$. It turns out that a better condition is to require that if in $v_1 \otimes \cdots \otimes v_n$ we have $v_i = v_j$ for some $i < j$ then its image in $\bigwedge^n(V)$ is zero. This implies the previous relation, but is a strictly stronger condition if $1 = -1$ in R. Thus, we define

$$J_n = \mathrm{Span}_R\{v_1 \otimes \cdots \otimes v_n : v_i \in V, \exists i < j \text{ s.t. } v_i = v_j\}.$$

Then $J = \oplus_{n=0}^{\infty} J_n$ is a graded two-sided ideal and we define the **exterior algebra**

$$\bigwedge^\bullet V = T^\bullet(V)/J;$$

It is an R-algebra. We denote the image of $v_1 \otimes \cdots \otimes v_n$ in $\bigwedge^\bullet V$ by

$$v_1 \wedge \cdots \wedge v_n.$$

If $R = k$ is a field, $e_1, \dots e_d$ is a basis for V, we can calculate that

$$\{e_{i_1} \wedge \cdots \wedge e_{i_n} : 1 \le i_1 < \cdots < i_n \le d\}$$

is a basis for $\bigwedge^n V$ and so

$$\dim_k(\bigwedge^n V) = \binom{d}{n}.$$

This is not obvious, but it is reasonable to leave it as an exercise. Some notable special cases are

$$\bigwedge^0 V = k, \quad \bigwedge^1 V = V, \quad \bigwedge^d V \cong k, \quad \bigwedge^n V = \{0\} \ \ \forall n > d.$$

Example 19.1.2. Decomposition of $V^{\otimes 2}$. Assume that V is a vector space over $\mathbb{C}$. The surjections

$$V \otimes V \to \mathrm{Sym}^2(V), \qquad V \otimes V \to \bigwedge^2 V,$$

have sections. For the image of the first section, we can take the subspace spanned by the vectors $\{\frac{1}{2}(e_i \otimes e_j + e_j \otimes e_i) : 1 \le i \le j \le d\}$, of dimension $d(d+1)/2$, and for the other the subspace spanned by $\{\frac{1}{2}(e_i \otimes e_j - e_j \otimes e_i) : 1 \le i < j \le d\}$ of dimension $d(d-1)/2$. Abusing notation, we denote them also

$$\mathrm{Sym}^2(V) \subset V, \qquad \bigwedge^2 V \subset V.$$

Note that $\bigwedge^2 V$ maps to zero under the projection $V \otimes V \to \mathrm{Sym}^2(V)$ and so we find that the two spaces are complementary and by dimension count

$$V \otimes V = \mathrm{Sym}^2(V) \oplus \bigwedge^2 V.$$

Passing to characters, and making use of Equation (23), we find the identity

$$\chi_V(g)^2 = \frac{1}{2}(\chi_V(g)^2 + \chi_V(g^2)) + \frac{1}{2}(\chi_V(g)^2 - \chi_V(g^2)),$$

where χ_V is the character of the representation V. We conclude that

$$\chi_{\bigwedge^2 V}(g) = \frac{1}{2}(\chi_V(g)^2 - \chi_V(g^2)).$$

Example 19.1.3. Let $\rho^{St,0}$ be the irreducible representation of dimension $n-1$ of S_n contained in its standard representation ρ^{St}. We claim that $\bigwedge^a \rho^{St,0}$ is irreducible as well for all $1 \le a \le n-1$. The proof here follows Fulton-Harris.

Lemma 19.1.4. *Let χ be the character of $\bigwedge^a \rho^{St}$ of S_n, for some $1 \leq a \leq n-1$. Then*

$$\|\chi\|^2 = 2.$$

Proof. Choose a basis $e_1, \ldots, e_n$ for ρ^{St}. Note that there is a bijection between subsets $B \subseteq \{1, 2, \ldots, n\}$ of a elements and basis elements of $\bigwedge^a \rho^{St}$; to $B = \{i_1 < \cdots < i_a\}$ is associated the basis element $e_{i_1} \wedge \cdots \wedge e_{i_a}$. Define for $\sigma \in S_n$ and such a subset B,

$$\sigma\{B\} = \begin{cases} 0 & \sigma(B) \neq B, \\ 1 & \sigma(B) = B,\ \sigma|_B \text{ is even}, \\ -1 & \sigma(B) = B,\ \sigma|_B \text{ is odd}. \end{cases}$$

The point of this definition is that

$$\chi(\sigma) = \sum_B \sigma\{B\}.$$

As this is a sum of integers

$$\bar{\chi}(\sigma) = \chi(\sigma).$$

Therefore,

$$\begin{aligned} \langle \chi, \chi \rangle &= \frac{1}{n!} \sum_{\sigma \in S_n} \Big(\sum_{|B|=a} \sigma\{B\} \Big)^2 \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} \sum_{|B|=a} \sum_{|C|=a} \sigma\{B\} \cdot \sigma\{C\} \\ &= \frac{1}{n!} \sum_{|B|=a} \sum_{|C|=a} \sum_{\substack{\sigma \in S_n \\ \sigma(B)=B \\ \sigma(C)=C}} \sigma\{B\} \cdot \sigma\{C\}. \end{aligned}$$

Now, any permutation σ such that $\sigma(B) = B$ and $\sigma(C) = C$ can be written as product,

$$\sigma = \sigma_1 \sigma_2 \sigma_3 \sigma_4,$$

where,

$$\sigma_1 \in S_{\{1,\ldots,n\} - \{B \cup C\}}, \quad \sigma_2 \in S_{B-C}, \quad \sigma_3 \in S_{C-B}, \quad \sigma_4 \in S_{B \cap C}.$$

Let also $\ell = \ell_{B,C} := \sharp B \cap C$. Below, we should note that the argument also works for $\ell = a$ under the conventions $S_\emptyset = \{1\}$ and $\mathrm{sgn}(1) = 1$. Then, continuing our calculation, we find that

$$\begin{aligned} \langle \chi, \chi \rangle &= \frac{1}{n!} \sum_{|B|=a} \sum_{|C|=a} \sum_{\sigma_1 \in S_{n-(2a-\ell)}} \sum_{\sigma_2 \in S_{a-\ell}} \sum_{\sigma_3 \in S_{a-\ell}} \sum_{\sigma_4 \in S_\ell} \mathrm{sgn}(\sigma_4)^2 \, \mathrm{sgn}(\sigma_2) \, \mathrm{sgn}(\sigma_3) \\ &= \frac{1}{n!} \sum_{|B|=a} \sum_{|C|=a} (n - (2a - \ell))! \cdot \ell! \times \Big(\sum_{\sigma_2 \in S_{a-\ell}} \mathrm{sgn}(\sigma_2) \Big)^2 \end{aligned}$$

In the calculation above we should note that ℓ is really $\ell_{B,C}$, i.e. it depends on B and C, but we omitted that from the notation which is quite cumbersome as it is. But, we should note now that if $a - \ell > 1$ then $\sum_{\sigma_2 \in S_{a-\ell}} \mathrm{sgn}(\sigma_2) = 0$ as this sum is $n! \cdot \langle \chi_{\mathrm{sgn}}, \mathbb{1} \rangle = 0$ for the sign representation and the trivial representation $\mathbb{1}$ of $S_{a-\ell}$. To continue the calculation we consider separately the case $\ell = a$ where $B = C$ and the case $\ell = a - 1$ where $B \cap C$ has $a - 1$ elements. We find, using that $\binom{n}{a}$ is the number of choices of B,

$$\begin{aligned} \langle \chi, \chi \rangle &= \frac{1}{n!} \left(\sum_{|B|=a} (n-a)!a! + \sum_{|B|=a} \sum_{\substack{|C|=a \\ |C \cap B|=a-1}} (n-a-1)!(a-1)! \right) \\ &= \frac{1}{n!} \left(\binom{n}{a} (n-a)!a! + \binom{n}{a} a(n-a) \cdot (n-a-1)!(a-1)! \right) \\ &= 2 \end{aligned}$$

□

Corollary 19.1.5. *The representations $\bigwedge^a \rho^{St,0}$ are irreducible for all $1 \leq a \leq n-1$.*

Proof. We may think of $\bigwedge^a \rho^{St,0}$ as a sub-representation of $\bigwedge^a \rho^{St}$. If

$$\bigwedge^a \rho^{St} = \oplus_i \rho_i^{a_i},$$

is the decomposition into irreducible representations then, by the Lemma,

$$2 = \|\chi\|^2 = \sum_i a_i^2.$$

There is only one possibility, that there are two irreducible representations in $\bigwedge^a \rho^{St}$ and they are non-isomorphic. But then, as $\bigwedge^a \rho^{St,0}$ is a proper sub-representation is must be equal to one of them. □

This is quite useful. For example, for S_5 we get this way 4 irreducible representations of dimensions $4, 6, 4, 1$ (these integers are the values $\dim(\wedge^a \rho^{St,0}) = \binom{4}{a}, a = 1, 2, 3, 4$). They are all distinct, which is clear except for the two 4-dimensional representations where it follows from calculating the characters. The 1-dimensional representation is the sign representation. We also have the trivial representation $\mathbb{1}$. As S_5 has $7 = p(5)$ irreducible representations and their dimensions satisfy $120 = 5! = 1^2 + 1^2 + 4^2 + 6^2 + 4^2 + x^2 + y^2$, we conclude that there are two additional irreducible representations, both of dimension 5.

19.2. **Tensors, wedges and multi-linear forms.** Let R be a commutative ring. Thus, every R-module is naturally an R-bimodule and so tensor products retain the property of being R-modules. Let V and W be R-modules. An R-d-**multilinear map** of V into W a function

$$f\colon V \times V \times \cdots \times V \to W, \quad (v_1, v_2, \ldots, v_d) \to f(v_1, v_2, \ldots, v_d),$$

which is R-linear in each variable separately. That is, for any $i = 1, \ldots, d$, $v_i, v_i' \in V$, $r, r' \in R$, we have

$$f(v_1, \cdots, rv_i + r'v_i', \ldots, v_d) = rf(v_1, \cdots, v_i, \ldots, v_d) + r'f(v_1, \cdots, v_i', \ldots, v_d).$$

The property

$$f(v_1, \cdots, rv_i, \ldots, v_d) = rf(v_1, \cdots, v_i, \ldots, v_d)$$

follows from that. A particular case is $d = 2$ giving us the notion of an R-bilinear pairing; note that this a different notion than R-biadditive function, although every R-bilinear pairing is also R-biadditive. Nonetheless, the d-fold tensor product $V^{\otimes d}$ over R serves the same function; see § 2.7

Lemma 19.2.1. *To give an R-d-multilinear map $V^d \to W$ is to give a homomorphism of R-modules*

$$V^{\otimes d} \to W.$$

Proof. Let f be a multilinear map then f is also R-multiadditive and by (a slight extension of) the results we have proven, we get a well-defined homomorphism of abelian groups

$$V^{\otimes d} \to W, \quad v_1 \otimes v_2 \otimes \cdots \otimes v_d \mapsto f(v_1, v_2, \ldots, v_d).$$

Now, $r \cdot v_1 \otimes v_2 \otimes \cdots \otimes v_d = (rv_1) \otimes v_2 \otimes \cdots \otimes v_d$ which is mapped to $f(rv_1, v_2, \ldots, v_d) = rf(v_1, v_2, \ldots, v_d)$. That shows that the map $V^{\otimes d} \to W$ is automatically R-linear. The converse is very similar. □

Let us use the notation

$$V^* := \mathrm{Hom}_R(V, R).$$

Some care is needed because in general $(V^*)^* \neq V$: although there is a natural map $V \to (V^*)^*$ taking $v \in V$ to the function $\varphi \mapsto \varphi(v)$, this map needs not be injective (or surjective). A good case to keep in mind is $R = \mathbb{Z}$, $V = \mathbb{Z}/2\mathbb{Z}$ where $V^* = \{0\}$ and so $V^{**} = \{0\}$ and the map $V \to V^{**}$ is not injective. In this notation, we find

$$\left\{\begin{array}{c}R\text{-}d\text{-multilinear maps}\\ \text{from } V \text{ to } W\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{homomorphisms of } R\text{-modules}\\ V^{\otimes d}\to W\end{array}\right\}$$

We say that an R-d-multilinear map f from V to W is **symmetric** if for any $\sigma \in S_d$ we have

$$f(v_{\sigma(1)},\dots,v_{\sigma(d)}) = f(v_1,\dots,v_d).$$

Equivalently, if the map $f\colon V^{\otimes d}\to W$ vanishes on the R-submodule spanned by $v_{\sigma(1)}\otimes\cdots\otimes v_{\sigma(d)} - v_1\otimes\cdots\otimes v_d$. We say that f is **antisymmetric** if for any $1\le i<j\le d$ we have

$$f(v_1,\dots,v_d) = 0 \text{ if } v_i = v_j.$$

Equivalently, if the map $f\colon V^{\otimes d}\to W$ vanishes on the R-submodule spanned by the tensors $v_1\otimes\cdots\otimes v_d$ with $v_i = v_j$. This implies that for any $\sigma\in S_d$ we have

$$f(v_{\sigma(1)},\dots,v_{\sigma(d)}) = \operatorname{sgn}(\sigma) f(v_1,\dots,v_d).$$

Corollary 19.2.2. *There are natural homomorphisms of R-modules*

$$\left\{\begin{array}{c}R\text{-}d\text{-multilinear symmetric maps}\\ \text{from } V \text{ to } W\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{homomorphisms of } R\text{-modules}\\ \operatorname{Sym}^d V\to W\end{array}\right\},$$

and

$$\left\{\begin{array}{c}R\text{-}d\text{-multilinear antisymmetric maps}\\ \text{from } V \text{ to } W\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{homomorphisms of } R\text{-modules}\\ \bigwedge^d V\to W\end{array}\right\}.$$

Corollary 19.2.3. *(Uniqueness of determinant) Let V be a free R module of rank d. Up to a scalar, there is a unique d-multilinear anti-symmetric map $V\times\cdots\times V\to R$ (i.e., a determinant map).*

Proof. Such maps correspond to homomorphisms of R modules

$$\bigwedge^d V\to R.$$

But $\bigwedge^d V\cong R$ and $\operatorname{Hom}_R(R,R) = R$. □

Remark 19.2.4. If we choose a basis for V we can represent an element of V by a column vector in R^d and we can represent an element of $V\times\cdots\times V$ (d-times) by a $d\times d$ matrix with entries in R. A d-multilinear antisymmetric map is then a function

$$M_d(R)\to R,$$

that is multilinear in the columns and changes sign when we switch the columns and (even stronger when 2 is not invertible in R) vanishes when two columns are the same. That is, this function has all the properties of the determinant and the Corollary asserts that it must be equal to the determinant up to a scalar.

19.2.1. *Existence of invariant forms.* Suppose that G is a finite group acting on a finite dimensional vector space (ρ, V). It is a natural question to ask if there is a non-zero G-invariant bilinear form on V.[25] Using the methods we developed, this is the same as asking if $(\operatorname{Sym}^2(V))^*$ has a G-invariant vector. Now, if χ denotes the character of (ρ,V), the character of $\operatorname{Sym}^2(V)$ is, by Equation (23),

$$\chi_{\operatorname{Sym}^2(V)}(g) = \frac{1}{2}(\chi_V(g)^2 + \chi_V(g^2)).$$

[25]We have asked that before for *hermitian* forms, saw that the answer is yes, and put it to good use in proving that every representation decomposes into a direct sum of irreducible representations in Theorem 10.2.2. Although one could address this question using the same methods we are using here for symmetric bilinear forms, it is a bit convoluted and so we will not do it here.

Thus, the character ψ of $(\mathrm{Sym}^2(V))^*$ is

$$\psi(g) = \frac{1}{2}(\overline{\chi_V(g)}^2 + \overline{\chi_V(g^2)}).$$

The dimension of the invariant subspace is $\frac{1}{|G|}\sum_{g\in G}\frac{1}{2}(\overline{\chi_V(g)}^2 + \overline{\chi_V(g^2)})$. Since this is an integer, we can take the complex conjugate and conclude:

Proposition 19.2.5. *V has a non-zero G-invariant symmetric bilinear form if and only if*

$$\frac{1}{|G|}\sum_{g\in G}\frac{1}{2}(\chi_V(g)^2 + \chi_V(g^2)) \neq 0.$$

More precisely, the dimension of the vector space of invariant symmetric bilinear forms is equal to $\frac{1}{|G|}\sum_{g\in G}\frac{1}{2}(\chi_V(g)^2 + \chi_V(g^2))$.

If V is an irreducible representation of G then using additional considerations one finds that there is at most one G-invariant bilinear form on V, up to scalar (symmetric or not). The argument is based on viewing a bilinear form $f(\cdot,\cdot)$ as a homomorphism $V \to V^*, v \mapsto f(v,\cdot)$ and G-equivariance is equivalent to this map being a homomorphism of representations; one uses now Schur's Lemma. We leave that as an exercise. Admitting this we conclude:

Corollary 19.2.6. *Let V be an irreducible representation of G. Then there is a symmetric bilinear G-invariant form on V if and only if $\frac{1}{|G|}\sum_{g\in G}\frac{1}{2}(\chi_V(g)^2 + \chi_V(g^2)) \neq 0$, in which case this form is unique up-to-scalar. This is the case if and only if χ is real-valued.*

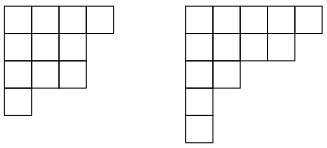
20. Representations of the symmetric group

The representations of the symmetric group are a rich area of research, with a very combinatorial flavour, as is to be expected. We will only provide a short introduction, very far from giving a true understanding of this topic.

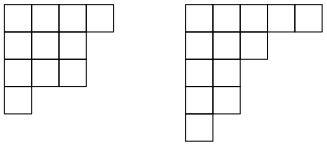
20.1. **Young tableuax.** Let $n \geq 1$ be an integer and λ be a partition of n that we write as

$$\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r),$$

where the λ_i are positive integers whose sum is n. To this partition we associate a **Young diagram** having λ_1 boxes in the first row, λ_2 boxes in the second row and so on. For example, to the partitions $(4,3,3,1)$ and $(5,4,2,1,1)$ we associate the diagrams



Conversely, a Young diagram defines a partition. If we transpose the Young diagram, making the columns into rows, we get the **conjugate partition** λ'. The conjugate diagrams to those above are, respectively,



corresponding to the conjugate partitions $(4,3,3,1)$ and $(5,3,2,2,1)$, respectively.

A **Young tableau** is obtained by numbering the boxes in the Young diagram using each of the numbers $\{1,2,\ldots,n\}$ exactly once. And by doing that we also create a permutation – its cycles are defined by the rows – with associated partition λ.

Here are two Young tableaux associated to the same Young diagram. The first one is called the **canonical** (or **standard**), one.

1	2	3
4	5	6
7	8	
9		

2	4	6
3	7	8
9	5	
1		

Given a tableau T there is an associated conjugate tableau T'. Even if T is standard, T' is not standard.

To a Young tableau we associate two subgroups of S_n. They depend on the tableau, not just on the underlying partition λ, but since they depend on the tableau up to conjugation only, we will usually be lax and write these subgroups as P_λ and Q_λ. If emphasizing the tableau T is important we will write P_T and Q_T. At any rate, define

$$P = P_\lambda = \{\sigma \in S_n : \sigma \text{ preserves every row of the tableau}\}$$

and

$$Q = Q_\lambda = \{\sigma \in S_n : \sigma \text{ preserves every column of the tableau}\}.$$

For example, for the tableau

1	2	3
4	5	6
7	8	
9		

we have

$$P = S_{\{1,2,3\}} \times S_{\{4,5,6\}} \times S_{\{7,8\}}, \quad Q = S_{\{1,4,7,9\}} \times S_{\{2,5,8\}} \times S_{\{3,6\}}.$$

For the tableau

2	4	6
3	7	8
9	5	
1		

we have

$$P = S_{\{2,4,6\}} \times S_{\{3,7,8\}} \times S_{\{5,9\}}, \qquad Q = S_{\{1,2,3,9\}} \times S_{\{4,5,7\}} \times S_{\{6,8\}}.$$

To a partition λ (or, rather, to a tableau) we associate two elements of the group ring $\mathbb{C}[S_n]$ of S_n:

$$a_\lambda = \sum_{\sigma \in P_\lambda} \sigma, \quad b_\lambda = \sum_{\sigma \in Q_\lambda} \operatorname{sgn}\sigma \cdot \sigma.$$

And we define the **Young symmetrizer** of λ as

$$c_\lambda = a_\lambda b_\lambda \in \mathbb{C}[S_n].$$

If we keep the Young diagram of λ but take another tableau associated to it, the elements $a_\lambda, b_\lambda, c_\lambda$ are instead $\tau a_\lambda \tau^{-1}, \tau b_\lambda \tau^{-1}, \tau c_\lambda \tau^{-1}$ for some $\tau \in S_n$ and any τ will arise this way for some tableau associated with λ.

Now, the point is that for *every* element y of a group ring $\mathbb{C}[G]$, the right ideal $\mathbb{C}[G]y$ is a left $\mathbb{C}[G]$-module. Thus, it defines a representation of the group G. In fact, the natural map

$$\mathbb{C}[G] \to \mathbb{C}[G]y, \qquad x \mapsto xy,$$

is a homomorphism of representations. Thus, by letting y take different values we get different quotient representations of the regular representation $\mathbb{C}[G]$.

20.2. **The irreducible representations V_λ.**

Theorem 20.2.1. *Let λ be a partition of n. Then*

$$V_\lambda := \mathbb{C}[S_n]\cdot c_\lambda$$

is an irreducible representation of S_n. Every irreducible representation of S_n is isomorphic to a representation obtained this way, for a unique partition λ.

Proof. We will need a few lemmas. We follow Fulton and Harris.

Lemma 20.2.2. *Let T be a tableau and $P = P_T, Q = Q_T$ be the associated stabilizers, $a = a_T = \sum_{p\in P} p, b = b_T = \sum_{q\in Q}\mathrm{sgn}(q)q, c = c_T = \sum_{(p,q)\in P\times Q}\mathrm{sgn}(q)pq$.*

1. *For $p\in P, pa = ap = a$ and $pc = c$.*
2. *For $q\in Q$, $qb = bq = \mathrm{sgn}(q)b$ and $cq = \mathrm{sgn}(q)c$.*
3. *For $p\in P, q\in Q$, $pcq = \mathrm{sgn}(q)c$ and, moreover, any element x of $\mathbb{C}[S_n]$ such that $pxq = \mathrm{sgn}(q)x$, for all $p\in P, q\in Q$, is a scalar multiple of c.*

Proof. The first two claims are rather clear given the explicit form of a and b. As $c = ab$ the assertion $pcq = \mathrm{sgn}(q)c$ is clear too. The main issue is to prove that this property characterizes c up to multiplication by a scalar. Let $x\in\mathbb{C}[G]$ be an element such that

$$pxq = \mathrm{sgn}(q)x, \qquad \forall p\in P, q\in Q.$$

Write $x = \sum_g n_g g$. The group G is a disjoint union of double cosets $G = \cup_{g_i} Pg_iQ$ and it is clear that the coefficient n_{g_i} determines the coefficient n_g for any $g\in Pg_iQ$. For example, if $g = pg_iq$ then, on the one hand, $pxq = \mathrm{sgn}(q)x = \sum_g \mathrm{sgn}(q)n_g g$ and, on the other hand, $pxq = \sum_g n_g pgq$ and we find that

$$n_g = \mathrm{sgn}(q)n_{g_i}, \text{ if } g = pg_iq. \tag{24}$$

For example, for $g = pq\in PQ$ we have

$$n_g = n_1\cdot\mathrm{sgn}(q).$$

Note that $P\cap Q = \{1\}$ as any permutation fixing every row and every column of a Young tableau fixes all its entries. (Conversely, *defining*, $n_g = n_1\cdot\mathrm{sgn}(q)$, and $n_g = 0$ for all $g\notin PQ$, is well-defined and gives us, in fact, $n_1\cdot c$.)

What remains to prove is that if $g\notin PQ$ then $n_g = 0$. If $g\notin PQ$ we will prove that $n_g = 0$ by finding a transposition τ such that $\tau\in P, g^{-1}\tau g\in Q$. Then, using Equation (24),

$$n_g = n_{\tau\cdot g\cdot g^{-1}\tau g} = \mathrm{sgn}(g^{-1}\tau g)n_g = -n_g,$$

and so

$$n_g = 0.$$

How can we find such a transposition? Let us be more precise and let T be the standard Young tableau associated to λ. The subgroups P, Q and the elements a, b, c are really P_T, Q_T, a_T, b_T, c_T. Let us consider the tableau gT obtained by applying g to every entry of T. We have

$$P_{gT} = gP_Tg^{-1}, \quad Q_{gT} = gQ_Tg^{-1}.$$

We claim that if $g\notin PQ$ then there is a pair of integers $i\neq j$ such that i, j appear in the same row of T and in the same column of gT. Given that, choose $\tau = (ij)$. Then, as $i, j,$ are in the same row of T, definitely $\tau\in P_T$. As i, j, are in the same column of gT, definitely $\tau\in Q_{gT}$ and so $g^{-1}\tau g\in Q_T$. Thus, we are done once we prove the following lemma:

Lemma 20.2.3. *Suppose that there is no pair of distinct integers that appear in the same row of T and the same column of gT then $g\in PQ$.*

Before proving the lemma, it may be a good illustration to look at a particular example. Consider the following Young tableau

$$T = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&5&6\\\hline 7&8\\\cline{1-2} 9\\\cline{1-1}\end{array}$$

We $P_T = S_{\{1,2,3\}} \times S_{\{4,5,6\}} \times S_{\{7,8\}}, Q_T = S_{\{1,4,7,9\}} \times S_{\{2,5,8\}} \times S_{\{3,6\}}$. Let us take $g = (15)(37)$. Then

$$gT = \begin{array}{|c|c|c|}\hline 5&2&7\\\hline 4&1&6\\\hline 3&8\\\cline{1-2} 9\\\cline{1-1}\end{array}$$

Then $1,2$ appear in the same row of T and same column of gT ($4,5$ are another example). On the other hand, let us take an element g that is clearly in P_TQ_T, say $g = (123)(78)(1479) = (14879231)$. Then,

$$gT = \begin{array}{|c|c|c|}\hline 4&3&1\\\hline 8&5&6\\\hline 9&7\\\cline{1-2} 2\\\cline{1-1}\end{array}$$

We can verify that there is no pair of integers that appears in the same row of T and the same column of gT. The lemma states that this, in turn, implies that $g \in PQ$.

Proof. (Lemma 20.2.3) We can find $p \in P$ and $q \in Q_{gT}$ such that pT and $q(gT)$ have the same first row. Indeed, find q by "raising to the top" in each column of gT the elements appearing in the first row of T and use p to rearrange the first row of T so that pT and $q(gT)$ have the same first row. This is possible to do because of the assumption that no two integers appear in the same row of T and the same column of gT – in particular, the integers in the first row of T are in different columns of gT and so in each column of gT there is precisely one integer from the first row of T.

In the example above, we can take $q = (9842)$ and $p = (123)$. Then

$$pT = \begin{array}{|c|c|c|}\hline 2&3&1\\\hline 4&5&6\\\hline 7&8\\\cline{1-2} 9\\\cline{1-1}\end{array} \qquad q(gT) = \begin{array}{|c|c|c|}\hline 2&3&1\\\hline 4&5&6\\\hline 8&7\\\cline{1-2} 9\\\cline{1-1}\end{array}$$

Now consider the tableaux $pT, q(gT)$, where we erase the first row. Their, P and Q, so to say, satisfy $P \subset P_{pT} = P_T, Q \subset qQ_{gT} = Q_{gT}$. And we can repeat the process getting a p', q' (in our example, $p' = Id, q' = Id$) and pass to third row to find p'', q'', and so on. In our example, we will then use $p'' = (78), q'' = Id$.

Thus, we can find $p \in P, q \in Q_{gT}$ such that $pT = q(gT)$. This implies that $p = qg$, or $pg^{-1}q^{-1} = 1$ and

$$g = p \cdot g^{-1}q^{-1}g \in P_T \cdot g^{-1}(Q_{gT})g = P_TQ_T.$$

□

□

Corollary 20.2.4. *For any $z \in \mathbb{C}[S_n]$ we have $c_\lambda z c_\lambda \in \mathbb{C}c_\lambda$. In particular, $c_\lambda^2 = n_\lambda c_\lambda$ for some $n_\lambda \in \mathbb{C}$.*

Proof. (Corollary) Let $x = c_\lambda z c_\lambda$. For $p \in P, q \in Q$, we have $pxq = (pc_\lambda)z(c_\lambda q) = \mathrm{sgn}(q)c_\lambda z c_\lambda = \mathrm{sgn}(q)x$. Thus, x is a scalar multiple of c_λ. □

We can now prove that $V_\lambda = \mathbb{C}[S_n]c_\lambda$ is an irreducible representation: It follows from the Corollary that $c_\lambda V_\lambda \subseteq \mathbb{C}c_\lambda$. Let W be a subrepresentation of V_λ and consider $c_\lambda W$. There are two possibilities.

- $c_\lambda W = \mathbb{C}c_\lambda$. This means that $c_\lambda = c_\lambda w$ for some $w \in W$. Then
$$V_\lambda = \mathbb{C}[S_n]c_\lambda = \mathbb{C}[S_n]c_\lambda w \subseteq \mathbb{C}[S_n]w \subseteq \mathbb{C}[S_n]W \subseteq W,$$
and so $W = V_\lambda$.

- $c_\lambda W = \{0\}$. In this case we want to show that $W = \{0\}$. We need a lemma.

  Lemma 20.2.5. *Let G be a finite group and let W be isomorphic to a subrepresentation of $\mathbb{C}[G]$. For example, any irreducible representation of G has this property. There exists an element $\varphi \in \mathbb{C}[G]$ such that $W \cong \mathbb{C}[G]\varphi$ as a $\mathbb{C}[G]$-module. Moreover, $\varphi^2 = \varphi$.*[26]

  Proof. We know that W is isomorphic to a subrepresentation of $\mathbb{C}[G]$, and so we may assume that W is already a subrepresentation of $\mathbb{C}[G]$. We can therefore decompose $\mathbb{C}[G]$, *as a $\mathbb{C}[G]$-module*, as

  $$\mathbb{C}[G] = W \oplus W^\perp,$$

  and let φ_W be the projection map $\mathbb{C}[G] \to W$, which is a map of $\mathbb{C}[G]$-modules. Now, consider the element $1 \in \mathbb{C}[G]$ and its decomposition

  $$1 = \varphi + \sigma.$$

  Obviously, $\varphi_W(1) = \varphi_W(\varphi) + \varphi_W(\sigma) = \varphi$. Therefore, $\varphi_W(\tau) = \tau\varphi_W(1) = \tau\varphi$. That is, the projection map φ_W is multiplication from the right by φ. Since φ_W is a projection, $\varphi_W^2 = \varphi_W$ and so $\varphi_W^2(1) = \varphi_W(1)$; equivalently, $\varphi^2 = \varphi$. □

  We return now to the case of the symmetric group and $W \subseteq V_\lambda$ such that $c_\lambda W = 0$. Then $W \cdot W \subseteq (\mathbb{C}[S_n]c_\lambda) \cdot W = \mathbb{C}[S_n] \cdot (c_\lambda W) = 0$. On the other hand, for φ as in Lemma 20.2.5, $\varphi = \varphi^2 \in W \cdot W$ and we get a contradiction, unless $\varphi = 0$. Namely, unless $W = \{0\}$.

This concludes the proof that V_λ is irreducible. We note several consequences of the considerations above:

- Taking the case $W = V_\lambda$, we get that $c_\lambda V_\lambda \neq \{0\}$ and so $c_\lambda V_\lambda = \mathbb{C}c_\lambda$.
- $c_\lambda^2 = n_\lambda c_\lambda$ for some $n_\lambda \in \mathbb{C}$ (already from Corollary 20.2.4).
- Consider the linear operator T, which is the multiplication from the right by c_λ:

  $$\mathbb{C}[S_n] \to \mathbb{C}[S_n], \qquad x \mapsto xc_\lambda.$$

  As

  $$c_\lambda = \sum_{(p,q)\in P\times Q} \mathrm{sgn}(q) \cdot pq,$$

  for every $\sigma \in S_n$ we have

  $$T(\sigma) = \sigma c_\lambda = \sum_{(p,q)\in P\times Q} \mathrm{sgn}(q) \cdot \sigma pq.$$

  From this we find that if we calculate the trace of T using the standard basis for $\mathbb{C}[S_n]$ then

  $$\mathrm{Tr}(T) = n!$$

  Now, the kernel of T, $\mathrm{Ker}(T)$, is a $\mathbb{C}[S_n]$-submodule of $\mathbb{C}[S_n]$ and the image of T, V_λ, is irreducible. If $V_\lambda \cap \mathrm{Ker}(T) \neq 0$ then in fact $V_\lambda \subseteq \mathrm{Ker}(T)$ and that implies $T^2 = 0$. But then the characteristic polynomial of T is $x^{n!}$ and in particular $\mathrm{Tr}(T) = 0$. Contradiction. Thus, $\mathrm{Ker}(T)$ doesn't intersect V_λ. On the other hand, the dimension of $V_\lambda = n! - \dim(\mathrm{Ker}(T))$. Therefore, there is a direct sum decomposition as $\mathbb{C}[S_n]$ modules:

  $$\mathbb{C}[S_n] = \mathrm{Ker}(T) \oplus V_\lambda.$$

  It follows that $\mathrm{Tr}(T) = \dim(V_\lambda) \cdot n_\lambda$ because T acts on V_λ by multiplication by n_λ. We conclude that $n_\lambda = n!/\dim(V_\lambda)$, a rational non-zero number.[27]

[26] An element r of a ring R that satisfies $r^2 = r$ is called an **idempotet**. Then $1 = (1-r) + r$ and $1 - r$ is also an idempotent. Moreover $(1-r)r = 0$ and it follows that $R = Rr \oplus R(1-r)$, as left R-modules. If R is commutative, each summand is a ring; r is the identity element of Rr and $(1-r)$ is the identity element of $R(1-r)$.

[27] In fact, an integer as the dimension of an irreducible representation divides the order of the group. We will not need this.

- Consequently, we find that
$$(\frac{1}{n_\lambda}c_\lambda)^2 = \frac{1}{n_\lambda}c_\lambda.$$
We conclude that in fact $\frac{1}{n_\lambda}c_\lambda$ is the idempotent defining V_λ, whose existence is guaranteed by Lemma 20.2.5.

To complete the proof of Theorem 20.2.1 we need to also prove that if $\lambda \neq \mu$ then $V_\lambda \ncong V_\mu$.

Write $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots), \mu = (\mu_1 \geq \mu_2 \geq \dots)$. We say that $\lambda > \mu$ if the first non-vanishing difference $\lambda_i - \mu_i > 0$. This provides a linear (lexicographic) order on S_n; for every $\lambda \neq \mu$ either $\lambda > \mu$ or $\mu > \lambda$.

Lemma 20.2.6. *If $\lambda < \mu$ then $c_\lambda \mathbb{C}[G] c_\mu = 0$ and in particular $c_\lambda c_\mu = 0$.*

Note that the lemma implies that $V_\lambda \ncong V_\mu$ for $\lambda \neq \mu$. Indeed, if they are isomorphic, we may assume $\lambda < \mu$ and then, on the one hand $c_\lambda V_\lambda \neq 0$, and on the other hand, $c_\lambda V_\lambda \cong c_\lambda \mathbb{C}[G] c_\mu = 0$. So it's enough to prove the lemma.

To prove the lemma, it suffices to prove that for all $g \in \mathbb{C}[S_n]$ we have $b_\lambda g a_\mu = 0$ and, in fact, it suffices to prove that for $g \in S_n$. It is thus enough to prove $b_\lambda \cdot g a_\mu g^{-1} = 0$. One way to think about it is that we can change the tableau T' used to construct a_μ from T' to gT'. So, it is enough to prove that
$$b_T a_{T'} = 0,$$
where T is the tableau used to define b_λ (that we can assume to be a standard tableau) and T' the tableau used to define a_μ, without assuming it to be standard.

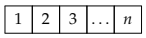
The assumption that $\lambda < \mu$ implies that there are two integers $i \neq j$ that are in the same column of λ and the same row of T'. To prove that we introduce some (non-standard) terminology. We will denote by $T[i]$ the i-th row of a tableau T and by $\mathrm{ctnt}(T[i])$ the numbers appearing in it.

The proof is by induction on n, where the case $n = 2$ is clear as there is only one possibility for $\lambda < \mu$ then, corresponding to the tableaux
$$T = \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline \end{array} \qquad T' = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline \end{array}$$
Consider the case $n > 2$, and denote $\lambda = (\lambda_1 \geq \dots \geq \lambda_r), \mu = (\mu_1 \geq \dots \geq \mu_s)$. If $\lambda_1 < \mu_1$ then $\mathrm{ctnt}(T'[1])$, that consists of μ_1 numbers, is distributed over the λ_1 columns of T and so there are $i \neq j$ in the first row of T' that are in the same column of T.

Suppose that $\lambda_1 = \mu_1$, and suppose that $\mathrm{ctnt}(T'[1])$ is distributed over all λ_1 columns of T (if this is not the case, then there are $i \neq j$ in the first row of T' that are in the same column of T and we are done). There exist then $q \in Q_T$ such that $\mathrm{ctnt}(T'[1]) = \mathrm{ctnt}(qT[1])$. Consider then the tableaux
$$qT^* = (qT[2], \dots, qT[r]), \qquad T'^* = (T'[2], \dots, T'[s]),$$
that are associated with the set $S^* = \{1, \dots, n\} \setminus \mathrm{ctnt}(T'[1])$, whose size n^* is smaller that n, and the partitions $\lambda^* = (\lambda_2 \geq \dots \geq \lambda_r), \mu^* = (\mu_2 \geq \dots \geq \mu_s)$. Note that $\lambda^* < \mu^*$, which implies $n^* \geq 2$. Applying the induction hypothesis, there are $i \neq j$ appearing in the same row of T'^* and the same column of qT^*. But that also means that i, j appear in the same column of T.

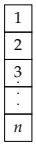
Thus, if these integers are i, j then $\tau = (ij) \in Q_T \cap P_{T'}$ and
$$b_T a_{T'} = (b_T \tau)(\tau a_{T'}) = -b_T a_{T'},$$
and we conclude $a_T b_{T'} = 0$. This concludes the proof of Theorem 20.2.1 □

Example 20.2.7. Consider the trivial Young tableau



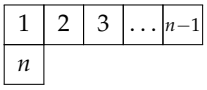
In this case $P = S_n, Q = \{1\}$ and so $a = \sum_{\sigma\in S_n}\sigma, b = 1$ and $c = \sum_{\sigma\in S_n}\sigma$. Note that for every σ we have $\sigma c = c$. Namely, $\mathbb{C}[S_n]\cdot c = \sum_{\sigma\in S_n}\mathbb{C}\sigma c = \mathbb{C}c$, with every σ acting trivially. That is, the associated representation is the trivial representation.

Example 20.2.8. Consider the Young tableau

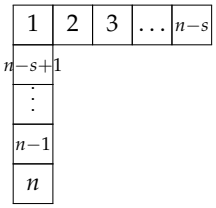


In this case $P = \{1\}, Q = S_n$ and so $a = 1, b = c = \sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\cdot\sigma$. Note that for every σ we have $\sigma c = \mathrm{sgn}(\sigma)c$. Therefore, $\mathbb{C}[S_n]\cdot c = \sum_{\sigma\in S_n}\mathbb{C}\sigma c = \mathbb{C}\,\mathrm{sgn}(\sigma)c = \mathbb{C}c$. Thus, the associated representation is just the 1-dimensional sign representation.

Example 20.2.9. A somewhat more complicated example is coming from the Young tableau



We claim that the associated representation is $\rho^{St,0}$. More generally, the representation associated to the tableau



is $\bigwedge^s\rho^{St,0}$. Let us sketch the argument for the case $s = 1$. In this case

$$P = S_{n-1},\quad Q = \langle(1n)\rangle,$$

and

$$a = \sum_{\sigma\in S_{n-1}}\sigma,\qquad b = 1-(1n),\qquad c = \sum_{\sigma\in S_{n-1}}\sigma - \sum_{\sigma\in S_{n-1}}\sigma\cdot(1n).$$

Now, quite generally, given a representation V of a group G, choose a vector $v\in V$, $v\neq 0$, and define

$$\mathbb{C}[G]\to V,\quad x\mapsto xv.$$

This is a morphism of $\mathbb{C}[G]$ modules. Given a left ideal I of $\mathbb{C}[G]$, I is a $\mathbb{C}[G]$ module and the map $\mathbb{C}[G]\to V$ induces a map of $\mathbb{C}[G]$-modules

$$I\to V,\quad x\mapsto xv.$$

If V is irreducible, the morphism $I\to V$ is either surjective, or zero. If I is irreducible, the map is either injective or zero.

Let us take the ideal $I = \mathbb{C}[S_n]c$ and the vector $v = (1,1,\ldots,-n-1)\in\rho^{St,0}$. Apply b to it: $bv = (n,0,\ldots,-n)$. Apply a to get $abv = (t,t,\ldots,t,-n\cdot(n-1)!)$, where $t = n\cdot(n-2)!$. In particular abv is not zero. Thus, we have a surjection

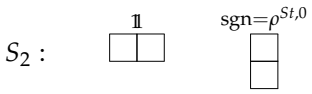
$$\mathbb{C}[G]c\twoheadrightarrow\rho^{St,0}.$$

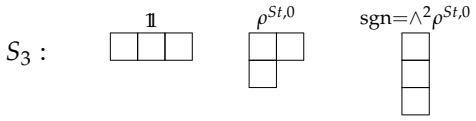
Since we know V_λ is irreducible, the map is an isomorphism.

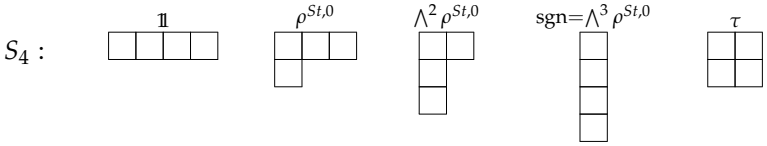
To show this is an isomorphism without using that V_λ is irreducible, it is enough to show that $\dim(\mathbb{C}[G]c)\leq n-1$. Using that for every $\sigma\in S_{n-1}$ we have $\sigma a = a$ and so $\sigma c = c$, we find that $\mathbb{C}[G]c = \sum_{i=1}^n\mathbb{C}\cdot(in)c$,

and then using that $(\sum_{i=1}^n (in))a = \sum_{\sigma\in S_n}\sigma$ we also find that $(\sum_{i=1}^n (in))c = 0$. It follows that $\dim(\mathbb{C}[G]c) \leq n-1$, as desired.

Assuming the result for $\bigwedge^s \rho^{st,0}$ we can match the Young diagrams with representations, at least for $n \leq 4$.





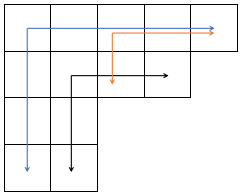


τ is the representation $\rho^{St,0}$ of S_3, pulled-back to S_4 through the homomorphism $S_4 \to S_3$.

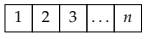
20.3. **Further results.** As one suspects, there is a lot of combinatorics involved in understanding what are the properties of the representations arising as $\mathbb{C}[G]c_\lambda$ (and some key words to google are "tabloid" and "Specht module"). For example, the dimension of the representation $V_\lambda = \mathbb{C}[G]c_\lambda$ is given by the **hook length formula**

$$\dim(V_\lambda) = \frac{n!}{\prod_{\text{all hooks}}(\text{hook length})}.$$

Every box in a Young diagram has a **hook** associated to it that consists off all boxes to the right of the box (and in the same row) and all boxes below the box (in the same column), including the initial box itself.



A diagram thus has n hooks. The **hook length** is just the number of the boxes in the hook. In the diagram above, 3 hooks are indicated and their lengths are 8, 5 and 4. The diagram



has n hooks, of lengths $n, n-1, \ldots, 1$. The dimension of the associated representation should therefore be 1 and we know that is correct. The diagram

has hooks of length $3,2,2,1$, giving a representation of dimension $4!/(3\times 2\times 2) = 2$.

A more general, and very natural, question is what is the character χ_λ of the representation V_λ? A formula for χ_λ will provide a formula for the dimension as $\dim(V_\lambda) = \chi_\lambda(1)$.

Let $g \in S_n$ be a permutation of cycle type $(i_1, \ldots, i_n)$ – meaning it has i_1 cycles of length 1, i_2 cycles of length 2, ..., i_n cycles of length n (and so $\sum_{j=1}^n j\cdot i_j = n$). Don't confuse that with the partition associated to g. Also use the notation $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k)$ for the partition λ. Define then

$$\ell_1 = \lambda_1 + k - 1, \ell_2 = \lambda_2 + k - 2, \ldots, \ell_k = \lambda_k.$$

And define polynomials $P_1,\dots,P_n$ in the variables $x_1,\dots,x_k$, by

$$P_j(x) = x_1^j + x_2^j + \cdots + x_k^j.$$

Let also,

$$\Delta(x) = \prod_{1\le i<j\le k}(x_i - x_j).$$

Frobenius' formula. *The character χ_λ is given by*

$$\chi_\lambda(g) = \textit{coefficient of } x_1^{\ell_1}\cdots x_k^{\ell_k} \textit{ in the polynomial } \Delta(x)\cdot\prod_{j=1}^n P_j(x)^{i_j}.$$

We will not prove this formula in this course. The proof can be found in Fulton & Harris.

21. Representations of $\mathrm{GL}_2(\mathbb{F})$, $\mathbb{F}$ a finite field.

To find all the complex representations of $\mathrm{GL}_2(\mathbb{F})$, where $\mathbb{F}$ is a finite field of $q = p^r$ elements, we begin by first determining the conjugacy classes. We will assume that $\mathbf{char}(\mathbb{F}) \neq 2$.

21.1. **Conjugacy classes in $\mathrm{GL}_2(\mathbb{F})$.** Recall a consequence the structure theorem for modules over PID. Matrices in $\mathrm{GL}_n(\mathbb{F})$ are classified up to conjugacy by their rational canonical form. For $\mathrm{GL}_2(\mathbb{F})$ the rational canonical forms corresponds to pairs of polynomials of the form $\{(x-a,x-a), a\in\mathbb{F}^\times\}$ of which there are $q-1$, to pairs of polynomials of the form $(x-a,x-b), a\neq b\in\mathbb{F}^\times$ of which there are $(q-1)(q-2)/2$, to polynomials of the form $(x-a)^2, a\in\mathbb{F}^\times$ of which there are $q-1$, and to quadratic irreducible polynomials $x^2+ax+b, a\in\mathbb{F}, b\neq 0\in\mathbb{F}$ of which there are $q(q-1)-(q-1)-(q-1)(q-2)/2 = q(q-1)/2$. These correspond to $\mathbb{F}[x]$-modules of the form

$$\mathbb{F}[x]/(x-a)\oplus\mathbb{F}[x]/(x-a), \qquad \mathbb{F}[x]/(x-a)\oplus\mathbb{F}[x]/(x-b), \qquad \mathbb{F}[x]/((x-a)^2)$$

and

$$\mathbb{F}[x]/(x^2+ax+b).$$

Examples of such matrices are provided by

$$\begin{pmatrix} a & 0\\ 0 & a\end{pmatrix}, \quad \begin{pmatrix} a & 0\\ 0 & b\end{pmatrix}, \quad \begin{pmatrix} a & 1\\ 0 & a\end{pmatrix}, \quad \begin{pmatrix} 0 & -b\\ 1 & -a\end{pmatrix}.$$

The centralizer of a matrix corresponds to the automorphism group of the module. Thus, in the first case the centralizer is $\mathrm{GL}_2(\mathbb{F})$, in the second case it is $\{\left(\begin{smallmatrix} x & 0\\ 0 & y\end{smallmatrix}\right) : x,y\in\mathbb{F}^\times\}$, in the third case it is $\{\left(\begin{smallmatrix} x & y\\ & x\end{smallmatrix}\right) : x\in\mathbb{F}^\times, y\in\mathbb{F}\}$ and in the fourth case it is the units of $\mathbb{F}[x]/(x^2+ax+b)$ which is a group of order q^2-1. This allows us, by the orbit-stabilizer formula, to determine how many elements are in each conjugacy class and we find the following:

rational form type	representative	no. of classes	size of conj. class
$(x-a,x-a)$	$\left(\begin{smallmatrix} a & 0\\ 0 & a\end{smallmatrix}\right)$	$q-1$	1
$(x-a,x-b)$	$\left(\begin{smallmatrix} a & 0\\ 0 & b\end{smallmatrix}\right)$	$(q-1)(q-2)/2$	$q(q+1)$
$(x-a)^2$	$\left(\begin{smallmatrix} a & 1\\ 0 & a\end{smallmatrix}\right)$	$q-1$	q^2-1
x^2+ax+b	$\left(\begin{smallmatrix} 0 & -b\\ 1 & -a\end{smallmatrix}\right)$	$q(q-1)/2$	q^2-q
		$=q^2-1$	

Thus, we must find q^2-1 distinct irreducible representations of $\mathrm{GL}_2(\mathbb{F})$. The simplest are the one dimensional characters. There are $q-1$ distinct homomorphisms $\alpha\colon \mathbb{F}^\times \to \mathbb{C}^\times$. Composing them with the determinant we find $q-1$ distinct one dimensional characters

$$\alpha(\det)\colon \mathrm{GL}_2(\mathbb{F}) \to \mathbb{C}^\times.$$

(In fact these are all the 1-dimensional characters. Equivalently, the commutator of $\mathrm{GL}_2(\mathbb{F})$ is $\mathrm{SL}_2(\mathbb{F})$, in fact for any field $\mathbb{F}$ and also for $\mathrm{GL}_n(\mathbb{F})$. But we will not need this fact.) We call the corresponding one dimensional representations U_α.

21.2. **Representations induced from a Borel.** Let B be the Borel subgroup of $\mathrm{GL}_2(\mathbb{F})$ that consists of the upper triangular matrices:

$$B = \left\{ \begin{pmatrix} a & x \\ 0 & b \end{pmatrix} : a,b \in \mathbb{F}^\times, x \in \mathbb{F} \right\}.$$

It is a subgroup of $\mathrm{GL}_2(\mathbb{F})$ of index $q+1$. The 1-dimensional characters of B are all of the form

$$\begin{pmatrix} a & x \\ 0 & b \end{pmatrix} \mapsto \alpha(a)\beta(b),$$

where α, β are characters of $\mathbb{F}^\times$. We denote these characters

$$\chi_{\alpha,\beta}\colon B \to \mathbb{F}^\times.$$

Define

$$W_{\alpha,\beta} = \mathrm{Ind}_B^{\mathrm{GL}_2(\mathbb{F})} \chi_{\alpha,\beta}.$$

It is a representation of $\mathrm{GL}_2(\mathbb{F})$ of dimension $q+1$.

Remark 21.2.1. Let G be a group, $H < G$ and (ρ, V) a representation of H. The induced representation was defined as

$$\mathrm{Ind}_H^G(\rho, V) = \mathbb{C}[G] \otimes_{\mathbb{C}[H]} V.$$

We now present another model for this representation.

Let $\mathrm{Rep} := \{g_i\}_{i=1}^d$ be left coset representatives for H. Then every element of the induced representation has the form $\sum_{i=1}^d g_i \otimes v_i$ for some $v_i \in V$, and this expression is unique. We can think of such an expression as a function $f\colon \mathrm{Rep} \to V$, such that $g_i \mapsto v_i$, and we extend it to a function $f\colon G \to V$ by

$$f(x) = f(g_i h_1) := \rho(h_1^{-1}) f(g_i) = \rho(h_1^{-1}) v_i, \quad x = g_i h_1, h_1 \in H.$$

This function satisfies

$$f(xh) = \rho(h^{-1}) f(x), \quad \forall x \in G, h \in H$$

(let $x = g_i h_1, h_1 \in H$ and calculate.) We temporarily define

$$ind(\rho, V) = \{f\colon G \to V | f(xh) = \rho(h^{-1}) f(x),\ \forall x \in G, h \in H\}$$

The process can be reversed and we find an isomorphism of *vector spaces*, $\mathrm{Ind}_H^G(\rho, V) \cong ind(\rho, V)$.

Via this construction, a pure tensor $g_i \otimes v_i$ gives a function f on G that is supported on the coset $g_i H$ and has the value $f(g_i h_1) = \rho(h_1)^{-1} v_i$. Let $g \in G$. It acts on $g_i \otimes v_i$ by sending it to $g g_i \otimes v_i = g_j \otimes \rho(h) v_i$, if $g g_i = g_j h$. This element, in turn, gives a function $G \to V$ supported on the coset $g_j H$ and having the value $\rho(h_1^{-1})\rho(h) v_i$ on $g_j h_1$.

On the other hand, if we let G act on functions $f\colon G \to V$ by translation $(gf)(x) = f(g^{-1}x)$, this action respects the subspace $ind(\rho, V)$ and for the function f we find $(gf)(g_j h_1) = f(g^{-1} g_j h_1) = f(g_i h^{-1} h_1) = \rho(h_1)^{-1}\rho(h) v_i$. We conclude an isomorphism of *representations*,

$$\mathrm{Ind}_H^G(\rho, V) \cong \{f\colon G \to V | f(xh) = \rho(h^{-1}) f(x),\ \forall x \in G, h \in H\},$$

where G acts on the right hand side by $(gf)(x) = f(g^{-1}x)$.

In particular, as in common in the literature, the induced representations from the Borel can be thought of as $\mathbb{C}$-valued functions on $\mathrm{GL}_2(\mathbb{F})$ that have an invariance property relative to the action of the Borel subgroup.

Lemma 21.2.2. *If $\alpha \neq \beta$, $W_{\alpha,\beta}$ is irreducible.*[28] *If $\alpha = \beta$ then $W_{\alpha,\alpha}$ is reducible and $W_{\alpha,\beta} = U_\alpha \oplus V_\alpha$, where V_α is an irreducible q-dimensional representation.*[29] *The representation $W_{\alpha,\beta}$ determines the unordered pair of characters $\{\alpha,\beta\}$ and the representation V_α determines the character α.*

Proof. We use Theorem 13.1.1 to calculate the induced character. For example, for an element of the form $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, there is a unique conjugate of it in B (i.e, itself) if $a = b$. And for $a \neq b$ it has $2q$ conjugates in B (the matrices $\begin{pmatrix} a & x \\ 0 & b \end{pmatrix}$, for any x, and the matrices $\begin{pmatrix} b & x \\ 0 & a \end{pmatrix}$, for any x). In the first case the size of the centralizer is $(q^2-1)(q^2-q)$ (the size of $\mathrm{GL}_2(\mathbb{F})$) and in the second case the size of the centralizer is $(q-1)^2$ for each of the cases. Thus, the value of the induced character is $(q+1)(\alpha(a)\beta(a))$ in the first case, and $\alpha(a)\beta(b)+\alpha(b)\beta(a)$ in the second case. We leave the calculation of the remaining cases as an exercise.

The character of $W_{\alpha,\beta}$ is given according to conjugacy classes as follows

$(x-a,x-a)$	$(x-a,x-b)$	$(x-a)^2$	x^2+ax+b
$(q+1)\alpha(a)\beta(a)$	$\alpha(a)\beta(b)+\beta(a)\alpha(b)$	$\alpha(a)\beta(a)$	0

Not that the character determines the pair $\{\alpha,\beta\}$. Now, as α,β take their values in roots unity, we have

$$\begin{aligned}\|\mathrm{Ind}\chi_{\alpha,\beta}\|^2 &= \frac{1}{(q^2-1)(q^2-q)}\Big(\sum_{a\in\mathbb{F}^\times}(q+1)^2+q(q+1)\sum_{\{a\neq b\}\subset\mathbb{F}^\times}(2+\alpha(a/b)\beta(b/a)+\alpha(b/a)\beta(a/b))+(q^2-1)\sum_{a\in\mathbb{F}^\times}1\Big)\\ &= \frac{1}{(q^2-1)(q^2-q)}[(q-1)(q+1)^2+q(q+1)(q-1)(q-2)+\\ &\qquad \frac{q(q+1)}{2}\sum_{\{(a,b):a\neq b,ab\neq 0\}}(\alpha(a/b)\beta(b/a)+\alpha(b/a)\beta(a/b))+(q^2-1)(q-1)].\end{aligned}$$

Now, use that for any group G and a non-trivial one dimensional character χ of G, $\sum_{g\in G}\chi(g)=0$, to deduce that if $\alpha\neq\beta$, then $\sum_{(a,b)}\alpha(a/b)\beta(b/a)=0$. We then get

$$\begin{aligned}\|\mathrm{Ind}\chi_{\alpha,\beta}\|^2 &= \frac{1}{(q^2-1)(q^2-q)}[(q-1)(q+1)^2+q(q+1)(q-1)(q-2)-\frac{q(q+1)}{2}\sum_{a\in\mathbb{F}^\times}2+(q^2-1)(q-1)]\\ &= \frac{1}{(q^2-1)(q^2-q)}[(q-1)(q+1)^2+q(q+1)(q-1)(q-2)-q(q+1)(q-1)+(q^2-1)(q-1)]\\ &= 1.\end{aligned}$$

Now, if $\alpha=\beta$ then

$$\begin{aligned}\|\mathrm{Ind}\chi_{\alpha,\beta}\|^2 &= \frac{1}{(q^2-1)(q^2-q)}[(q-1)(q+1)^2+q(q+1)(q-1)(q-2)+2q(q+1)(q-1)(q-2)+(q^2-1)(q-1)]\\ &= 2.\end{aligned}$$

It follows that $W_{\alpha,\alpha}$ is a sum of two non-isomorphic irreducible representations. We claim that the representation $\alpha\circ\det$ appears in $W_{\alpha,\alpha}$. To show that we can use Frobenius reciprocity

$$\mathrm{Hom}_G(W_{\alpha,\alpha},\alpha\circ\det)\cong\mathrm{Hom}_H(\chi_{\alpha,\alpha},\alpha\circ\det|_H).$$

[28] Those might be called principal series representations, although the name is typically used for p-adic groups of Lie groups (where a normalization factor is introduced).

[29] We might call V_α a **Steinberg** representation.

But $\chi_{\alpha,\alpha} = \alpha \circ \det|_H$, so the r.h.s. is non-zero. We conclude that

$$W_{\alpha,\alpha} = \alpha(\det) \oplus V_\alpha,$$

where V_α is irreducible of dimension q with character

$(x-a, x-a)$	$(x-a, x-b)$	$(x-a)^2$	x^2+ax+b
$q\alpha(a^2)$	$\alpha(ab)$	0	$-\alpha(b)$

Note that the character determines α. □

So far, we found the following distinct irreducible representations:

(1) $q-1$ representations $\alpha(\det)$ of dimension 1.
(2) $(q-1)(q-2)/2$ representations of dimension $q+1$.
(3) $q-1$ representations of dimension q.

We are therefore missing $q^2-1-(q-1+(q-1)(q-2)/2+q-1) = q(q-1)/2$ irreducible representations. If they are all the same dimension that dimension should be $q-1$.

Let $\epsilon \in F^\times$ be a non-square. Then $F[x]/(x^2-\epsilon)$ is a field with q^2 elements that we denote L. Writing

$$L = F \oplus F\sqrt{\epsilon},$$

multiplication becomes an F-linear transformation. $a \in F$ acts by the diagonal matrix $\mathbf{diag}(a,a)$ and $b\sqrt{\epsilon}$ for $b \in F$ acts by the matrix $\left(\begin{smallmatrix} 0 & b\epsilon \\ b & 0 \end{smallmatrix}\right)$. We get a homomorphism

$$L^\times \to \mathrm{GL}_2(F), \qquad \zeta := a + b\sqrt{\epsilon} \mapsto \left(\begin{smallmatrix} a & b\epsilon \\ b & a \end{smallmatrix}\right).$$

Following Fulton and Harris we will call the image K, but think of ζ also as an element of K. The elements in K such that $b \neq 0$ are elements of L that are not in F. The characteristic polynomial of such elements and its discriminant Δ are the following

$$t^2 - 2at + (a^2 - b^2\epsilon), \qquad \Delta = 4b^2\epsilon.$$

As ϵ is not a square, this is an irreducible polynomial. We observe that we get all the irreducible quadratic monic polynomials in $F[x]$ this way. Therefore, the $(q^2-q)/2$ pairs of elements of K, $\zeta = a \pm b\sqrt{\epsilon}$ for which $b \neq 0$ are representatives to the conjugacy classes of matrices with irreducible characteristic polynomial.

The group K is a cyclic group and we can take any of its q^2-1 characters $\varphi\colon K \to \mathbb{C}^\times$. When we look at $\mathrm{Ind}_K^G \varphi$ we get a q^2-q-dimensional representation whose character is given as follows.

$(x-a, x-a)$	$(x-a, x-b)$	$(x-a)^2$	$\zeta = a \pm b\sqrt{\epsilon}$
$q(q-1)\varphi(a)$	0	0	$\varphi(\zeta) + \varphi(\zeta^q)$

The argument is easy: the only elements of K with reducible characteristic polynomial are the diagonal matrices $\mathbf{diag}(a,a), a \in F^\times$. The only conjugates of $\zeta = a + b\sqrt{\epsilon}$ that lie in K are $a \pm b\sqrt{\epsilon}$.

At this point, we are going to pull a rabbit out of a hat. Suppose that $\varphi \neq \varphi^q$ and consider the class function χ^φ (a notation we choose so as to show the dependence on φ, yet to distinguish it from the character of φ, which is φ, and from the character of $\mathrm{Ind}_K^G \varphi$).

$$\chi^\varphi = \chi_{V_1 \otimes W_{\alpha,1}} - \chi_{W_{\alpha,1}} - \chi_{\mathrm{Ind}_K^G \varphi}.$$

It has the following values on conjugacy classes

Also $\|\chi^\varphi\|^2 = 1$ and $\chi^\varphi(1) = q-1 > 0$. If we write φ as an $\mathbb{Z}$-linear combination of the irreducible characters, possibly with negative coefficients, we can deduce that in fact φ is an irreducible character, corresponding to an irreducible representation X_φ. Note that φ and φ^q give the same character, but that

$(x-a,x-a)$	$(x-a,x-b)$	$(x-a)^2$	$\zeta=a\pm b\sqrt{\epsilon}$
$(q-1)\varphi(a)$	0	$-\varphi(a)$	$-\varphi(\zeta)-\varphi(\zeta^q)$

is the extent of the identifications. There are $q-1$ characters φ such that $\varphi=\varphi^q$ and $(q^2-1)-(q-1)$ such that $\varphi\neq\varphi^q$. Thus, we get $(q^2-q)/2$ irreducible representations of dimension $q-1$. By looking at the character tables we can recognize that all the irreducible representations we have constructed are non-isomorphic. Therefore, we found all the complex irreducible representations for $\mathrm{GL}_2(F)$, where F is a finite field.

22. Introduction to Fourier analysis on finite groups.

In this section we are following the wonderful book by P. Diaconis, *"Group representations in probability and statistics"* and if you find the following sections interesting, I very much recommend reading it; you should have essentially all the prerequisite knowledge for that. Another good and friendly reference is the book by A. Terras, *"Fourier Analysis on Finite Groups and Applications"*.

Before commencing, let us mention that the theory of Fourier transform for groups has many applications to other branches of science (computer science, chemistry, physics, electrical engineering), and even within mathematics to many branches besides probability and statistics; for example, to number theory.

The theory developed below resembles classical Fourier analysis, either for periodic functions on the real line $f(x)=f(x+1)$, or just for any function on the real line. In the first case one uses the functions $\{e^{2\pi in\cdot x}=\cos(2\pi n\cdot x)+i\sin(2\pi n\cdot x): n\in\mathbb{Z}\}$, and in the second case the functions $\{e^{2\pi ir\cdot x}: r\in\mathbb{R}\}$. In the first case, these functions are continuous unitary homomorphisms from $\mathbb{R}/\mathbb{Z}$ to $\mathbb{C}$ and in the second case continuous unitary homomorphisms $\mathbb{R}\to\mathbb{C}$.

There several possible conventions when defining the Fourier transform and the L^2-norm and these choices lead to some differences between the formulas appearing below in the case of finite groups and the classical formulas appearing in Fourier analysis. But once the different normalizations are reconciled, this is really the same theory done for various groups: $\mathbb{R}/\mathbb{Z},\mathbb{R}$ and finite groups G.

22.1. **Convolution.** Let G be a finite group. Let

$$C(G,\mathbb{C})=\{f: G\to\mathbb{C}\},$$

be the vector space of complex-valued functions on G. It is of course just the vector space $\mathbb{C}[G]$ we used many times before, but we change our notation as to emphasize the function-theoretic aspects. A function f defines an element $\sum_g f(g)[g]$ of $\mathbb{C}[G]$, and conversely. The complex vector space $C(G,\mathbb{C})$ has dimension $\sharp G$.

For $g\in G$ define the **delta function** $\delta_g: G\to\mathbb{C}$ by

$$\delta_g(x)=\begin{cases}1, & g=x;\\ 0, & \text{else.}\end{cases}$$

This function corresponds to the group-ring element $[g]\in\mathbb{C}[G]$. The collection $\{\delta_g: g\in G\}$ is a basis for $C(G,\mathbb{C})$.The group action in terms of functions is

$$(gf)(h):=f(g^{-1}h),\qquad g\in G, f\in C(G,\mathbb{C}).$$

Indeed, to see that, it is enough to consider the δ-functions δ_h associated to $[h]\in\mathbb{C}[G]$. For $g\in G$, the function $x\mapsto(g\delta_h)(x)=\delta_h(g^{-1}x)$ is the δ-function at gh, a function that corresponds to the element $[gh]\in\mathbb{C}[G]$.

We define the **convolution** of two functions $f, g \in C(G, \mathbb{C})$ as

$$(f * g)(x) = \sum_{s \in G} f(xs^{-1})g(s).$$

Note that for a non-abelian group in general $f * g \neq g * f$. In fact, convolution is just the product in the ring $\mathbb{C}[G]$; if we write an element of $\mathbb{C}[G]$ as $\sum_g a_g[g]$, where $a_g \in \mathbb{C}$, then

$$(\sum_g a_g[g]) + (\sum_g b_g[g]) = \sum_g (a_g + b_g)[g], \qquad (\sum_g a_g[g])(\sum_g b_g[g]) = \sum_g (\sum_s a_{gs^{-1}} b_s)[g].$$

And so it is clear that $C(G, \mathbb{C})$ is a ring under addition of functions and convolution, with identity element δ_1. For the same reason, the following two properties are evident, nonetheless we prove the first in the language of convolutions.

- $\delta_g * \delta_h = \delta_{gh}$.
- $f = \sum_g f(g)\delta_g$.

Indeed, $(\delta_g * \delta_h)(x) = \sum_{s \in G} \delta_g(xs^{-1})\delta_h(s) = \delta_g(xh^{-1})$, which is a function that is everywhere zero except at $x = gh$ where it is 1. Thus, $\delta_g * \delta_h = \delta_{gh}$.

22.2. **The Fourier transform.** The **Fourier transform** $\hat{f}$ of a function $f \in C(G, \mathbb{C})$ is a function on representations (ρ, V) of G. It associates to a representation ρ the element

$$\hat{f}(\rho) = \sum_{s \in G} f(s)\rho(s) \in \mathrm{End}(V).$$

In this part of the course, we will always assume that the representations are unitary, which we can always achieve by a suitable inner-product (cf. the proof of Maschke's theorem, Theorem 10.2.2).

Lemma 22.2.1. *We have the following properties of the Fourier transform:*

(1) $\widehat{f + g} = \hat{f} + \hat{g}$, *and* $\widehat{\alpha f} = \alpha \hat{f}$, $\alpha \in \mathbb{C}$.
(2) $\hat{\delta_g}(\rho) = \rho(g)$.
(3) Let U *be the* **uniform distribution** *on* G, $U(g) = \frac{1}{|G|}, \forall g \in G$. *Let* (ρ, V) *be a representation of* G. *Then* $\hat{U}(\rho)$ *is the projection operator on the sub-representation* V^G. *Thus, if* ρ *is irreducible and* $\rho \ncong \mathbb{1}$ *then* $\hat{U}(\rho) = 0$, *while* $\hat{U}(\rho)(\mathbb{1}) = Id$.
(4) Letting $T \cdot S$ *denote composition of linear transformations (the product in* $\mathrm{End}(V)$*) we have*

$$\widehat{f * g} = \hat{f} \cdot \hat{g} \tag{25}$$

Proof. The first two properties are immediate from the definition. The third property is just the definition of the projection operator and the fact that V^G is a subrepresentation of V. For the last, let (ρ, V) be a representation of G. Then,

$$\begin{aligned} \widehat{f * g}(\rho) &= \sum_{s \in G} (\sum_{t \in G} f(st^{-1})g(t)) \cdot \rho(s) \\ &= (\sum_{x \in G} f(x)\rho(x))(\sum_{t \in G} g(t)\rho(t)) \\ &= \hat{f}(\rho) \cdot \hat{g}(\rho). \end{aligned}$$

(The last product means product in the ring $\mathrm{End}(V)$, which is composition of linear functions.) □

22.3. **Fourier Inversion and Plancherel's formula.** The following theorem is very much reminiscent of formulas in Fourier analysis over $\mathbb{R}$.

Theorem 22.3.1. *Let $\rho_1,\dots,\rho_h$ be unitary representatives for the irreducible representations of G and let $d_i=\dim(\rho_i)$, $\chi_i=\chi_{\rho_i}$.*

(1) (Fourier Inversion) For any function $f\in C(G,\mathbb{C})$,

$$f(s)=\frac{1}{|G|}\sum_{i=1}^{h} d_i\cdot \mathrm{Tr}(\rho_i(s^{-1})\hat{f}(\rho_i)). \tag{26}$$

(2) (Plancherel's formula) For any two functions $f_1,f_2\in C(G,\mathbb{C})$,

$$\sum_{s\in G} f_1(s^{-1})f_2(s)=\frac{1}{|G|}\sum_{i=1}^{h} d_i\cdot \mathrm{Tr}(\hat{f}_1(\rho_i)\hat{f}_2(\rho_i)). \tag{27}$$

Proof. The proof is surprisingly simple for such scary looking formulas. First note that by linearity and bilinearity, it is enough to prove Fourier inversion for the functions δ_g, and the Plancherel formula for the pair of functions δ_g,δ_y. We first verify Fourier inversion for δ_g. In this case, the right hand side of (26) evaluated at s is:

$$\begin{aligned}\frac{1}{|G|}\sum_{i=1}^{h} d_i\cdot \mathrm{Tr}(\rho_i(s^{-1})\widehat{\delta_g}(\rho_i)) &= \frac{1}{|G|}\sum_{i=1}^{h} d_i\cdot \mathrm{Tr}(\rho_i(s^{-1})\rho_i(g))\\ &= \frac{1}{|G|}\sum_{i=1}^{h} d_i\cdot \mathrm{Tr}(\rho_i(s^{-1}g))\\ &= \frac{1}{|G|}\sum_{i=1}^{h} d_i\chi_i(s^{-1}g)\\ &= \frac{1}{|G|}\chi^{reg}(s^{-1}g).\end{aligned}$$

This is a function that vanished everywhere, except at $s=g$, where it receives the value 1. Namely, this is just the function $\delta_g(s)$, as required.

Now for the second part of the theorem. The right-hand side of Plancherel's formula (27) is equal to

$$\begin{aligned}\frac{1}{|G|}\sum_{i=1}^{h} d_i\mathrm{Tr}(\hat{\delta}_g(\rho_i)\hat{\delta}_y(\rho_i)) &= \frac{1}{|G|}\sum_{i=1}^{h} d_i\mathrm{Tr}(\rho_i(g)\rho_i(y))\\ &= \frac{1}{|G|}\sum_{i=1}^{h} d_i\mathrm{Tr}(\rho_i(gy))\\ &= \frac{1}{|G|}\sum_{i=1}^{h} d_i\chi_i(gy)\\ &= \frac{1}{|G|}\chi^{reg}(gy).\end{aligned}$$

This expression is equal to 1 if $g=y^{-1}$, and is equal to 0 otherwise. The sum

$$\sum_{s\in G}\delta_g(s^{-1})\delta_y(s)$$

has exactly the same property, so we get the equality we were after. □

We now derive a variant of Plancherel's formula that is very useful for applications. Recall the (potentially confusing, but customary) notation for a complex matrix M: $M^*=\bar{M}^t$. Recall also that if V is an inner product space we can define an inner product on $\mathrm{End}(V)$ by $\langle S,T\rangle=\mathrm{Tr}(S^*\cdot T)$ (it would be conjugate linear in the first variable and linear in the second variable), where S^* is the adjoint operator relative to the inner

product on V. If we choose an orthonormal basis for V then we may think about S as a matrix and S^* has the same meaning as above. This sheds some light on the expressions appearing in the next corollary.

Corollary 22.3.2. *Let f_1 be a real-valued function then, for any function f_2,*

$$\sum_{s\in G} f_1(s)f_2(s) = \frac{1}{|G|}\sum_{i=1}^{h} d_i \cdot \mathrm{Tr}((\hat{f}_1(\rho_i))^* \cdot \hat{f}_2(\rho_i)). \tag{28}$$

Proof. Let g be the function $g(s) = f_1(s^{-1})$. Then $\sum_{s\in G} f_1(s)f_2(s) = \sum_{s\in G} g(s^{-1})f_2(s)$ and we can apply Plancherel's formula to this sum. It only remains to note that for $\rho = \rho_i$ for some i,

$$\hat{g}(\rho) = \sum_s f_1(s^{-1})\rho(s) = \sum_s f_1(s)\rho(s^{-1}) = \sum_s f_1(s)\rho(s)^* = (\sum_s f_1(s)\rho(s))^* = \hat{f}_1(\rho)^*,$$

where we used that ρ_i is unitary and f_1 is real-valued. □

22.4. **The case of cyclic groups.** Let N be a positive integer and G the group $\mathbb{Z}/N\mathbb{Z}$. Any finite cyclic group is isomorphic to $\mathbb{Z}/N\mathbb{Z}$ and the isomorphism is canonical up to choice of a generator. Besides the groups $\mathbb{Z}/N\mathbb{Z}$, other cyclic groups that appear very often are the cyclic groups $\mathbb{F}^\times$, where $\mathbb{F}$ is a finite field. In the latter case, there is no canonical choice of a generator.

If G is an abelian group then $G \cong G^*$, where $G^* = \mathrm{Hom}(G, \mathbb{C}^\times)$ is its character group. The isomorphism requires choices in general, but in the case of $\mathbb{Z}/n\mathbb{Z}$ we can make it canonical by

$$\mathbb{Z}/N\mathbb{Z} \cong \mathrm{Hom}(\mathbb{Z}/N\mathbb{Z}, \mathbb{C}^\times), \quad a \mapsto \chi_a,$$

where

$$\chi_a(x) = e^{\frac{2\pi i a}{N}x} = \zeta_N^{ax},$$

where $\zeta_N = e^{2\pi i/N}$. Note that unlike our usual conventions, $\chi_0 = \mathbb{1}$ and not χ_1.

When G is abelian, every irreducible representation is an element of G^* and as the Fourier transform is additive relative to direct sum of representations, we see that the whole story depends on the Fourier transform evaluated on irreducible representations. Fix an isomorphism $G \cong G^*$. With this understanding, we can think about the Fourier transform for an abelian group G as a linear transformation

$$C(G, \mathbb{C}) \to C(G, \mathbb{C}), \qquad f \mapsto \hat{f}.$$

When $G = \mathbb{Z}/N\mathbb{Z}$ this is the map taking f to $\hat{f}(a) := \hat{f}(\chi_a) = \sum_{s=0}^{N-1} f(s)\chi_a(s)$.[30] That is,

$$\hat{f}(a) = \sum_{s=0}^{N-1} f(s)e^{2\pi i a s/N}$$

Our calculation of $\widehat{\delta_g}$ as the function sending a representation ρ to $\rho(g)$ gives that $\widehat{\delta_g}(a) = e^{2\pi i a g/N} = \chi_g(a)$. That is, the Fourier transform sends the function δ_g to the character χ_g:

$$\widehat{\delta_g} = \chi_g$$

The Fourier inversion formula $f(s) = \frac{1}{|G|}\sum_{i=1}^{h} d_i \cdot \mathrm{Tr}(\rho_i(s^{-1})\hat{f}(\rho_i))$ translates to the formula $f(s) = \frac{1}{n}\sum_{a=0}^{N-1} \chi_a(s^{-1})\hat{f}(a)$. Considering that $\chi_a(s^{-1}) = \bar{\chi}_a(s)$, we get[31]

$$f = \tfrac{1}{N}\textstyle\sum_{a=0}^{N-1} \hat{f}(a)\bar{\chi}_a = \tfrac{1}{N}\sum_{a=0}^{N-1} \hat{f}(a)\chi_{-a}$$

[30] Terras, who follows more closely the classical conventions, define $\hat{f}(a) = \hat{f}(\chi_{-a})$ leading to $\hat{f}(a) = \sum_{s=0}^{N-1} f(s)e^{-2\pi i a s/N}$.

[31] With Terras's conventions, one gets $f = \frac{1}{N}\sum_{a=0}^{N-1} \hat{f}(a)\chi_a$.

Moreover, from Plancherel's formula (or perhaps more simply by repeating its arguments) one can deduce, for any abelian group G of cardinality N in fact, that

$$\langle f,g\rangle = \tfrac{1}{N}\langle \hat{f},\hat{g}\rangle, \qquad \|f\|^2 = \tfrac{1}{N}\|\hat{f}\|^2 \tag{29}$$

22.5. **The uncertainty principle.** One version of the uncertainty principle states that one cannot pinpoint too-well both a function and its Fourier transform. If function f is simple, in the sense that it is concentrated around a small set (supported on a small set, for finite groups) then many oscillations (characters for finite groups) are required to describe it – meaning $\hat{f}$ is not concentrated around a small set.

The terminology originates with uncertainty principle in quantum mechanics. The latter can be phrased in terms of properties of the real Heisenberg group. We will not discuss how this is done, but focus on some mathematical aspects. One of many references to consult, one where the mathematical aspects are perhaps clarified, is the book by G. B. Folland/*Harmonic analysis in phase space*.

22.5.1. *The real Heisenberg group.* Given a $\mathbb{C}$-valued function $f \in L^2(\mathbb{R})$, there are two ways $g,h \in \mathbb{R}$ can act on such a function – "multiplication" and translation:

$$(M_g f)(x) := e^{2\pi i g x} f(x), \qquad (T_h f)(x) := f(x-h).$$

These operators do not commute. In fact,

$$T_h M_g = e^{-2\pi i g h} M_g T_h$$

(and so $T_{h_1} M_{g_1} T_{h_2} M_{g_2} = e^{2\pi i g_1 h_2} T_{h_1+h_2} M_{g_1+g_2}$). This non-commutativity is a manifestation of the uncertainty principle in Physics.

Define the real **Heisenberg group** $\mathcal{H} = \mathcal{H}(\mathbb{R})$ by

$$\mathcal{H} = \mathbb{R}\times\mathbb{R}\times\mathbb{R}, \qquad (\alpha_1,h_1,g_1)(\alpha_2,h_2,g_2) = (\alpha_1+\alpha_2+g_1h_2, h_1+h_2, g_1+g_2).$$

By observing that this is a semi-direct product of the form $(\mathbb{C}\times\mathbb{R})\rtimes\mathbb{R}$, or by the association $(\alpha,h,g) \leftrightarrow \begin{pmatrix} 1 & g & \alpha \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix}$, one sees that this is indeed a group. It acts on functions $f\colon \mathbb{R}\to\mathbb{C}$ by

$$(\alpha,h,g)(f) = e^{2\pi i\alpha} T_h(M_g f).$$

The centre $Z(\mathcal{H})$ of $\mathcal{H}$ is $\mathbb{C} = \{(\alpha,0,0) : \alpha\in\mathbb{R}\}$. One can define a pairing on $\mathbb{R}^2 = \mathcal{H}/Z(\mathcal{H})$ by:

$$\langle (h_1,g_1),(h_2,g_2)\rangle := [(0,h_1,g_1),(0,h_2,g_2)],$$

where on the right hand side we are taking the commutator (and our notation is $[x,y] = xyx^{-1}y^{-1}$). It is an element that lies in $Z(\mathcal{H})$ and identifying it with $\mathbb{R}$ we find that

$$\langle\cdot,\cdot\rangle : \mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}, \quad \langle (h_1,g_1),(h_2,g_2)\rangle = g_1h_2 - g_2h_1.$$

A key fact about the Heisenberg group is the classification of its representations.

Theorem 22.5.1 (Stone-von Neumann)**.** *$\mathcal{H}$ has a unique continuous unitary irreducible Hilbert space representation in which the centre acts through the character $\alpha\mapsto e^{2\pi i\alpha}$. This representation is, up to isomorphism, the action of $\mathcal{H}$ on $L^2(\mathbb{R})$.*

We remark that it is easy to generalize everything we said thus far to $\mathbb{R}^n$ and the Stone-von Neumann theorem still holds. In fact, it holds in much greater generality than for groups of the form $\mathbb{R}^n$ and for any non-trivial unitary character $Z(\mathcal{H})$, not just for the character $\alpha\mapsto e^{2\pi i\alpha}$.

22.5.2. *Models of the irreducible representation.* Besides the model of $L^2(\mathbb{R})$ already discussed, one can construct another model for this irreducible representation. The representation space is still $L^2(\mathbb{R})$, but here we let the Heisenberg group $\mathcal{H}$ act by

$$(\alpha, h, g)f = e^{2\pi i\alpha} M_{-h} T_g f.$$

As $M_{-h_1} T_{g_1} M_{-h_2} T_{g_2} = e^{2\pi i g_1 h_2} M_{-(h_1+h_2)} T_{g_1+g_2}$, this is another representation of $\mathcal{H}$ in which the center acts by $\alpha \mapsto e^{2\pi i\alpha}$ and is, likewise, irreducible. Thus, there should be an automorphism of $L^2(\mathbb{R})$, say $f \mapsto \tilde{f}$, the commutes with the two group actions. In fact, by Schur's Lemma, such an isomorphism is *unique, up to a scalar*. Namely, reducing the requirement to generators $(0, g, 0), (0, 0, h)$, we require:

$$\widetilde{M_g f} = T_g \tilde{f}, \quad \widetilde{T_h f} = M_{-h} \tilde{f}.$$

The Fourier transform for functions on the real line is defined as[32]

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(s) e^{-2\pi i s\xi} ds.$$

Fourier inversion is

$$f(s) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i s\xi} d\xi,$$

and the normalizations are such that we have

$$\langle f_1, f_2 \rangle = \langle \hat{f}_1, \hat{f}_2 \rangle.$$

The Fourier transform satisfies

$$\widehat{(e^{2\pi i g(\cdot)} f)}(\xi) = \hat{f}(\xi - g) = T_g \hat{f}(\xi), \qquad \widehat{T_h f}(\xi) = e^{-2\pi i h\xi} \hat{f}(\xi) = M_{-h} \hat{f}(\xi).$$

Namely, the Fourier transform is an equivariant isomorphism between two models of the same irreducible representation; a so-called **intertwining operator**.

As said, the uncertainty principle in quantum mechanics has very much to do with the canonical unitary representation of the (suitably generalized) Heisenberg group and the non-commutativity of the operators M_g and T_h. The precise explanation is beyond the scope of these notes. In the context of our discussion, the uncertainty principle implies that one cannot at the same time localize the value of a function and its Fourier transform. To be precise, one has the inequality for a function f of norm 1:

$$\text{The uncertainty principle :} \qquad \left(\int_{-\infty}^{\infty} (x-a)^2 |f(x)|^2 \, dx \right) \left(\int_{-\infty}^{\infty} (\xi - b)^2 |\hat{f}(\xi)|^2 \, d\xi \right) \geq \frac{1}{16\pi^2}.$$

Thus, the above says that if f is concentrated around $a \in \mathbb{R}$ then no matter what $b \in \mathbb{R}$ is chosen, $\hat{f}$ cannot be concentrated around b. Equality is attained in the above if and only if f is, modulo translation and multiplication by a phase factor, a Gaussian function (i.e. of the form Ke^{-cx2}).

22.5.3. *Finite Heisenberg groups.* Let G be a finite abelian group with N elements. Define the **finite Heisenberg group** as

$$\mathcal{H} = \mu_N \times G \times G^*, \quad (\alpha_1, h_1, g_1)(\alpha_2, h_2, g_2) = (\alpha_1\alpha_2 \cdot g_1(h_2), h_1 h_2, g_1 g_2).$$

Here μ_N are the N-roots of unity. A straightforward verification shows this is a group with identity $(1, 1, 1)$ and inverses $(\alpha, h, g)^{-1} = (\alpha^{-1} g(h), h^{-1}, g^{-1})$. The centre is $Z(\mathcal{H}) = \{(\alpha, 0, 0) : \alpha \in \mu_N\}$, and the commutator pairing is

$$\langle \cdot, \cdot \rangle \colon (G \times G^*) \times (G \times G^*), \quad \langle (h_1, g_1), (h_2, g_2) = [(1, h_1, g_1), (1, h_2, g_2)] = g_1(h_2) g_2(h_1)^{-1}.$$

The analogue of the Stone-von Neumann theorem is

[32] Here we adhere to the usual convention.

Theorem 22.5.2. *$\mathcal{H}$ has a unique irreducible representation in which the centre acts by*

$$(\alpha,1,1)\mapsto \alpha\cdot Id.$$

Proof. The proof is very similar to the proof of Blichfeldt's theorem 18.3.1, so we will be brief. Let

$$\underline{G}=\{(1,h,1):h\in G\},\quad \underline{G}^*=\{(1,1,g):g\in G^*\}.$$

Both are subgroups of $\mathcal{H}$ and the projections $\underline{G}\to G, \underline{G}^*\to G^*$ are isomorphisms.

Let (ρ,V) be an irreducible representation in which the centre acts as scalars. We will write $\rho(h)=\rho(1,h,1), \rho(g)=\rho(1,1,g)$. As $\underline{G}$ is abelian, there is a decomposition

$$V=\oplus_{g\in G^*}V_g,\quad V_g:=\{v\in V:\rho(h)(v)=g(h)\cdot v,\forall h\in\underline{G}\}.$$

An easy computation shows that if $v\in V_g$ and $g_1\in G^*$ then $\rho(g_1)v\in V_{g_1^{-1}g}$. From this one can deduce the following:

- all the V_g have the same dimension;
- choose some non-zero $v_1\in V_1$ and let $v_g=\rho(g)v_1$ then $v_g\in V_{g^{-1}}$ and $\rho(g_1)v_g=v_{g_1g}$;
- $V=\oplus_{g\in G^*}\mathbb{C}\cdot v_{g^{-1}},\quad \mathbb{C}\cdot v_{g^{-1}}=V_g$.

(It is in the last point that we use irreducibility.) Note that then (ρ,V) is uniquely determined because we know how $\mu_N,\underline{G},\underline{G}^*$ act on each v_g.[33] □

22.5.4. *Models for the irreducible representation.* We construct two models of the unique irreducible representation of Theorem 22.5.2.

First model: We take as our vector space $V=L^2(G)=C(G,\mathbb{C})$. An element $h\in G$ acts on functions by translation

$$(T_hf)(x)=f(h^{-1}x).$$

An element $g\in G^*$ acts on functions by "multiplication"

$$(M_gf)(x)=g(x)f(x).$$

An easy verification shows the identities:

- $M_{g_1}M_{g_2}=M_{g_1g_2},\ T_{h_1}T_{h_2}=T_{h_1h_2}$.
- $T_hM_g=g(h)^{-1}M_gT_h$.
- $T_{h_1}M_{g_1}T_{h_2}M_{g_2}=g_1(h_2)T_{h_1h_2}M_{g_1g_2}$.

Consequently, we have a group action of $\mathcal{H}$ on V by

$$(\alpha,h,g)f=\alpha\cdot(T_h(M_g(f))).$$

Namely, $(\alpha,h,g)\mapsto\alpha\cdot T_h\circ M_g$. By comparing dimensions we see that V must be isomorphic to the unique irreducible representation of $\mathcal{H}$ such that the centre acts naturally.

Second model: We take as our vector space $V_1=L^2(G^*)=C(G^*,\mathbb{C})$. An element $h\in G$ acts on functions "by multiplication"

$$(M_hf)(\xi)=\xi(h)f(\xi).$$

An element $g\in G^*$ acts on functions by translation

$$(T_gf)(\xi)=f(g^{-1}\xi).$$

[33] In fact, we have proven that $(\rho,V)\cong\mathrm{Ind}_{\mu_N\times\underline{G}}^{\mathcal{H}}\mathbb{1}$. To see that, take as representatives for the cosets of $\mu_N\times\underline{G}$ the elements of $\underline{G}^*$. Then $\mathrm{Ind}_{\mu_N\times\underline{G}}^{\mathcal{H}}=\oplus_{g\in\underline{G}^*}\mathbb{C}\cdot g$. If we let $v_g=[g]$ we find that $g_1\in\underline{G}^*$ acts as $g_1v_g=v_{g_1g}$. As $h\in G$ satisfies $hg=(1,h,1)(1,1,g)=(1,h,g)=(1,1,g)(g(h)^{-1},h,1)=g(h)^{-1}gh$, it follows that h in G acts on v_g by multiplication by the scalar $g^{-1}(h)$. That is, $\mathbb{C}\cdot g$ is precisely the eigenspace associated to g^{-1}.

An easy verification shows the identity

$$T_g M_h = g^{-1}(h) M_h T_g, \quad g \in G^*, h \in G.$$

Consequently, we have a group action of $\mathcal{H}$ on V_1 by

$$(\alpha, h, g) f = \alpha \cdot (M_h(T_{g^{-1}}(f))).$$

Namely, $(\alpha, h, g) \mapsto \alpha \cdot M_h \circ T_{g^{-1}}$. By dimension considerations we see that V_1 must be isomorphic to the unique irreducible representation of $\mathcal{H}$ such that the centre acts naturally.

We conclude that there is an isomorphism of representations of $\mathcal{H}$, *unique up to a scalar*,

$$L^2(G) \to L^2(G^*).$$

If we denote this isomorphism $f \mapsto \hat{f}$ then it must satisfy, in particular, that

$$\widehat{T_h f} = M_h \hat{f}, \quad \widehat{M_g f} = T_{g^{-1}} \hat{f},$$

and those properties, in turn, guarantee that $f \mapsto \hat{f}$ is equivariant for the action of $\mathcal{H}$. The following Lemma shows that the Fourier transform satisfies these identities and hence our notation is justified. The isomorphism is, indeed, the Fourier transform!

Lemma 22.5.3. *Let G be a finite abelian group. The Fourier transform satisfies*

$$\widehat{T_h f} = M_h \hat{f}, \quad \widehat{M_g f} = T_{g^{-1}} \hat{f}.$$

Proof. The first identity is

$$\widehat{T_h f}(\xi) = \sum_{s \in G} (T_h f)(s) \xi(s) = \sum_{s \in G} f(h^{-1}s) \xi(s) = \sum_{s \in G} f(s) \xi(hs) =$$
$$\xi(h) \sum_{s \in G} f(s) \xi(s) = (M_h \hat{f})(\xi).$$

The second is,

$$\widehat{M_g f}(\xi) = \sum_{s \in G} (M_g f)(s) \xi(s) = \sum_{s \in G} g(s) f(s) \xi(s) = \sum_{s \in G} f(s) (g\xi)(s) = (T_{g^{-1}} \hat{f})(\xi).$$

□

22.5.5. *The uncertainty principle for finite groups.*

Theorem 22.5.4. *Let G be a finite abelian group and let f be a non-zero function on G with Fourier transform $\hat{f}$ on G^*. Then*

$$|\mathrm{supp}(f)| \cdot |\mathrm{supp}(\hat{f})| \geq |G|.$$

Example 22.5.5. If $f = \delta_g$ then $|\mathrm{supp}(f)| = 1$, so $\hat{f}$ should be a no-where vanishing function. Indeed, we saw that $\hat{f} = \chi_g$ which takes values in the unit circle and in particular is no-where vanishing.

Proof. Unlike many of our results, this is not a statement that can be reduced by linearity to delta functions. We will estimate $\|f\|^2$ using the Fourier transform and the identity (29).

On the one hand,

$$\|f\|^2 = \frac{1}{|G|} \sum_{s \in G} |f(s)|^2 \leq \frac{1}{|G|} |\mathrm{supp}(f)| \cdot \max\{|f(x)|^2\}. \tag{30}$$

On the other hand, by Fourier inversion,

$$f(x) = \frac{1}{|G|} \sum_{\chi \in G^*} \hat{f}(\chi) \bar{\chi}(x).$$

As each $|\chi(x)|=1$, we find that

$$\sup\{|f(x)|^2\} \leq \frac{1}{|G|^2}(\sum_{\chi\in G^*}|\hat{f}(\chi)|)^2 \leq \frac{1}{|G|^2}(\sum_{\chi\in G^*}|\hat{f}(\chi)|^2)\cdot|\mathrm{supp}(\hat{f})|.$$

(using Cauchy-Schwartz for the sum $(\sum_{\chi\in G^*}|\hat{f}(\chi)|\cdot 1_{\mathrm{supp}(\hat{f})})^2$). Using (29), we find

$$\max\{|f(x)|^2\} \leq \|f\|^2\cdot|\mathrm{supp}(\hat{f})|.$$

Combining with (30), and then dividing by $\|f\|^2$, gives the theorem. □

22.6. **Random walks on cyclic groups.** Let N be a positive integer, and consider the integers modulo N, $\mathbb{Z}/N\mathbb{Z}$. For various applications in cryptography, statistics, computer science and more, it is of interest to randomly choose a congruence class modulo N, or to emulate a random walk on $\mathbb{Z}/N\mathbb{Z}$. True randomness is hard; it's hard to generate and hard to "excavate" from nature. For that reason, one tries to expand, or stretch, a small amount of randomness to create a process that is pseudo-random; it is not completely random, but for all practical purposes it is.

Consider then the following process

$$x_{k+1}=a_kx_k+b_k,\quad k=1,2,\dots.$$

At each iteration a_k and b_k can be chosen among the classes $(\mathbb{Z}/N\mathbb{Z})^\times$ and $\mathbb{Z}/N\mathbb{Z}$, respectively, according to some agreed upon distribution. (This process is related to pseudo-random number generators, but we will not get into that here.) The simplest situation that is not completely deterministic is

$$a_k=1,\forall k,\quad b_k \text{ chosen from } \{\pm 1\} \text{ with equal probability.}$$

This process just requires a fair coin-toss at every step.

Let us denote functions on $\mathbb{Z}/N\mathbb{Z}$ by vectors $(a_0,\dots,a_{N-1})$. And let us suppose that the initial seed is $x_0=0\in\mathbb{Z}/N\mathbb{Z}$, namely, it is the vector $(1,0,\dots,0)$ with probability 1. Then, the distribution after one iteration is $P=(0,1/2,0,\dots,0,1/2)$, and after n-steps it is given by $P^{*n}:=P*P*\cdots*P$ (convolution n-times). For example, applying the random walk twice, it is clear that we can only end at $0,2$ or $-2=N-2$, and the probability we end at 0 can be found as

$$P(b_1=1)\cdot P(b_2=-1)+P(b_1=-1)\cdot P(b_2=1)=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}.$$

Similarly the probability for ending at 2 is $P(b_1=1)\cdot P(b_2=1)=1/4$, and so on. We recognize that we are just calculating $P*P$. For example, $(P*P)(0)=\sum_{j=0}^{p-1}P(j)P(-j)=P(1)P(p-1)+P(p-1)P(1)=1/2$.

Let us switch for a moment to multiplicative notation (which will hopefully be less confusing), and write $\mathbb{Z}/N\mathbb{Z}=\langle t\rangle$ where $t^N=1$. Using the group-ring notation (instead of the functions on G notation), we can say that

$$P=\frac{1}{2}(t+\frac{1}{t}),$$

and so

$$P^{*n}=\frac{1}{2^n}(t+\frac{1}{t})^n=\frac{1}{2^n}\sum_{j=0}^{n}a_j(n)t^j,$$

where

$$a_j(n)=\sum_{i\in\{0,\dots,n\},2i-n\equiv j(N)}\binom{n}{i}.$$

The limiting distribution is thus

$$\lim_{n\to\infty}P^{*n}=\lim_{n\to\infty}(a_0(n),a_1(n),\dots,a_{N-1}(n)).$$

Our main interest is to know whether $\lim_{n\to\infty} P^{*n}$ approaches the uniform distribution U, and, if so, how fast? If N is even, P^{*n} does not approach the uniform distribution as it is always support on the even integers modulo N. Suppose henceforth that N is odd. The fact that P^{*n} approaches the uniform distribution U is fairly easy (and follows from basic theory of Markov chains). The main question is *how quickly* it approaches U.

To gauge this we introduce the **total variation norm** $\|\cdot\|_{max}$. Let G be a finite group. For any two probability distributions $P, Q \in C(G, \mathbb{C})$ we let

$$\|P-Q\|_{max} = \max_{A\subset G} |P(A)-Q(A)| = \frac{1}{2}\sum_{g\in G} |P(g)-Q(g)|,$$

where $P(A) = \sum_{a\in A} P(a)$ is the probability of the event A.[34]

Lemma 22.6.1 (Diaconis-Shahshahani)**.** *Let G be a finite group with irreducible (unitary) representations $\rho_1 = \mathbb{1}, \ldots, \rho_h$. and let P be a probability distribution on G. Then, using $*$ to denote the adjoint operator,*

$$\|P-U\|_{max}^2 \leq \frac{1}{4}\sum_{i=2}^{h} d_i \cdot \mathrm{Tr}(\hat{P}(\rho_i)^* \cdot \hat{P}(\rho_i)).$$

(Namely, the trivial representation $\mathbb{1}$ is the only one not appearing in this sum.)

We will prove this lemma later on. Let us first see its application for the process we are discussing. In this case, recall that the irreducible representations of $\mathbb{Z}/N\mathbb{Z}$ are the 1-dimensional representations $\{\chi_j : j = 0, 1, \ldots, N-1\}$, where

$$\chi_j(a) = \zeta_N^{aj} \quad (\zeta_N = e^{2\pi i/N}).$$

(Namely, χ_j is the character such that $\chi_j(1)$ is the N-th root of unity $e^{2\pi ij/N}$.) Then,

$$\hat{P}(\chi_j) = \frac{1}{2}(\chi_j(1) + \chi_j(-1)) = \cos(2\pi j/N).$$

By multiplicativity of the Fourier transform, using $*n$ in the the exponent to note n-fold convolution,

$$\widehat{P^{*n}}(\chi_j) = \cos(2\pi j/N)^n.$$

Applying the Diaconis-Shahshahani lemma we find

$$\|P^{*n}-U\|_{max}^2 \leq \frac{1}{4}\sum_{j=1}^{N-1} \cos(2\pi j/N)^{2n}.$$

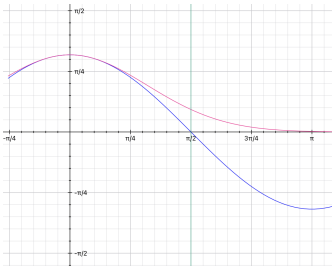
This last sum, though elementary in appearance, is not that easy to estimate, yet a relatively simple argument[35] gives a bound and one gets the following. If $N \geq 7$ and *odd* then for $n \geq N^2$

$$\|P^{*n}-U\|_{max}^2 \leq e^{-\frac{\pi^2}{2}\cdot\frac{n}{N^2}}.$$

This can be formulated qualitatively as saying that

[34] The total variation norm is just half the L^1 norm.

[35] One proves that for $x \in [0, \pi]$ we have $\cos(x) \leq e^{-x^2/2}$ and then sums the resulting geometric series. For more details see Diaconis' book, p. 26.



"for $a_k \equiv 1$, and b_k chosen uniformly from the set $\{1,-1\}$, about N^2 iterations of the process

$$x_{k+1} = a_k x_k + b_k$$

are required to achieve a distribution close to the uniform distribution."

One can perform a similar analysis for the case $a_k = 1$ and b_k chosen uniformly from $\{0,1,-1\}$ and get a very similar result. On the other hand, in stark-contrast, one can prove the following results for N such that $\gcd(N,6) = 1$:

"for $a_k \equiv 3$, and b_k chosen uniformly from $\{1,-1\}$, about $\log N$ iterations of the process $x_{k+1} = a_k x_k + b_k$ are required to achieve a distribution close to the uniform distribution."

It is quite surprising that the second type of random walk mixes so well compared to the first type. One reason the estimates are so different is that we are transitioning from representation theory for the group $\mathbb{Z}/N\mathbb{Z}$ to representation theory for the group $\mathbb{Z}/N\mathbb{Z} \rtimes \mathbb{Z}/N\mathbb{Z}^\times$. The process $x_{k+1} = 3x_k + b_k$ is thought of as coming from a random walk on the group $\mathbb{Z}/N\mathbb{Z} \rtimes \mathbb{Z}/N\mathbb{Z}^\times$ corresponding to taking powers of the random element $(b,3)$, where $b = \{1,0,-1\}$ with equal probability.

22.7. **Proof of the Diaconis-Shahshahani lemma.** Let us now prove the lemma. Recall the statement:

Let G be a finite group with irreducible (unitary) representations $\rho_1 = \mathbb{1}, \dots, \rho_h$. and let P be a probability distribution on G then

$$\|P - U\|_{max}^2 \leq \frac{1}{4} \sum_{i=2}^{h} d_i \mathrm{Tr}(\hat{P}(\rho_i)^* \cdot \hat{P}(\rho_i)).$$

(Namely, the trivial representation $\mathbb{1}$ is the only one not *appearing in this sum.)*

Proof. Applying the Cauchy-Schwarts inequality for real numbers $(\sum a_n b_n)^2 \leq (\sum a_n^2)(\sum b_n^2)$ and taking all the $b_n = 1$, we find that

$$4\|P - U\|_{max}^2 = (\sum_{s \in G} |P(s) - U(s)|)^2 \leq \sharp G \cdot \sum_{s \in G} (P(s) - U(s))^2.$$

We view the last sum as $\sum_{s \in G} f(s)h(s)$, where $f(s) = h(s) = (P(s) - U(s))$. Apply the version of Plancherel's formula given in Corollary 22.3.2 to find

$$\sharp G \cdot \sum_{s \in G} (P(s) - U(s))^2 \leq \sum_{i=1}^{h} d_i \mathrm{Tr}(\hat{f}(\rho)^* \cdot \hat{f}(\rho))$$

Now, $\hat{f}(\rho_i) = (\hat{P} - \hat{U})(\rho_i)$ and, using Lemma 22.2.1, we see that it is equal to $\hat{P}(\rho_i)$ for $\rho_i \neq \mathbb{1}$ (i.e., for $i > 1$), while $\hat{f}(\mathbb{1}) = (\hat{P} - \hat{U})(\mathbb{1}) = 1 - 1 = 0$. Therefore, we find

$$\sum_{i=1}^{h} d_i \mathrm{Tr}(\hat{f}(\rho)^* \cdot \hat{f}(\rho)) = \sum_{i=2}^{h} d_i \mathrm{Tr}(\hat{P}(\rho)^* \cdot \hat{P}(\rho)),$$

and the proof is complete. □

22.8. **Random walks on Cayley graphs.** [36] Let G be a group and S a set of elements of G. We construct a directed graph whose vertices are the elements of G and where for every $s \in S$ and vertex $g \in G$, there is an edge between g and sg. We will assume that S is symmetric: $s \in S \Leftrightarrow s^{-1} \in S$. Under this condition there is also an edge from sg to g as $g = s^{-1}(sg)$. We may therefore form an *undirected graph* that we denote $\Gamma(G,S)$ and call the **Cayley graph** of G with respect to S. It has a vertex set G and there is an edge between the vertices g_1 to g_2 if for some $s \in S$, $sg_1 = g_2$. Note that $\Gamma(G,S)$ is has no parallel edges and if

[36] A good reference for this section is *Hoory, Linial and Wigderson: Expander graphs and their applications. Bull. Amer. Math. Soc. 43 (2006), no. 4, 439–561.*

$1 \notin A$, it has no loops. Moreover, it is a regular graph of degree $d = |S|$. It is connected if and only if S is a generating set of G.

The group G acts as (right) automorphisms of $\Gamma(G,S)$ where $g \in G$ sends a vertex g_1 to $g_1 g$. If there is an edge between g_1 and g_2, there is an edge between $g_1 g$ and $g_2 g$ and so we indeed get an automorphism of the graph.

Assume that G is finite and enumerate the elements of G as $G = \{g_1, g_2, \ldots, g_N\}$, where $N = |G|$. Let S be a symmetric generating set of G such that $1 \notin S$. We define the **normalized adjacency matrix** A of G as the matrix whose ij entry is equal to $\frac{1}{d}$ if there is an edge between g_1 and g_2 and is 0 otherwise. The matrix A is symmetric, because S is symmetric, has 0's on the diagonal because $1 \notin S$ and is bi-stochastic, meaning it has non-negative entries and the sum of every row and every column is 1. We may therefore also view A as defining a bi-stochastic **Markov chain** on the set of vertices. The bi-stochastic property and the assumption that S is a generating set implies, by basic results in Markov chains, the following:

- that there is a unique probability distribution U such that $AU = U$;
- that U is the uniform distribution: $U = {}^t(\frac{1}{N}, \ldots, \frac{1}{N})$;
- that for any probability distribution P we have $\lim_{n\to\infty} A^n P = U$.

Thus, starting from any vertex g (a probability distribution concentrated at the vertex g) performing the random walk A, which means that at every step we move to a neighbour, where the choice of any particular neighbour is with probability $1/d$, the random process converges to the uniform distribution. That means, that after enough iterations of the random process the chances of being at any particular vertex g' are more or less independent of the vertex.

The main interest, thus, is to get a quantitive estimate on the rate of convergence in

$$\lim_{n\to\infty} A^n P = U.$$

We have already dealt with case of a cyclic group $\mathbb{Z}/N\mathbb{Z}$ and $S = \{1, -1\}$. We would likewise want to use group representations to study the general case.

With the notation and assumptions above (G, N, S, d, etc.), given any representation ρ of G we define

$$A_\rho = \frac{1}{d} \sum_{s\in S} \rho(s).$$

We may think about A_ρ as the Fourier transform of the (normalized) characteristic function $\frac{1}{d}\mathbb{1}_S$ of the generating set S. We claim that the adjacency matrix A is obtained this way from the regular representation ρ^{reg} of G:

$$A = A_{\rho^{reg}}.$$

Indeed, taking the delta functions $\{[g_i]\}$ as a basis for the regular representation, an element $s \in S$ acts by

$$\rho(s)[g_i] = [sg_i],$$

and in term of a matrix it has an ij entry that is zero, unless $g_j = sg_i$.

To continue the analysis it will convenient to assume that $\Gamma(G,S)$ is not a bipartite graph. Namely, there is no partition of the vertices into two non-empty sets such that all edges go from one set to the other. This is no always the case; for example, if S is the set of transpositions in S_n then the graph is bipartite with one set being the even permutations and the other the odd. For any G, If S has an element of odd order then $\Gamma(G,S)$ is not bipartite. So our assumption allows for great many examples.

If $\Gamma(G,S)$ is not bipartite then one can show the following (by simple linear algebra considerations):

- 1 is an eigenvalue of multiplicity 1 of A with eigenvector $(\frac{1}{N}, \ldots, \frac{1}{N})$;

- any other eigenvalue λ of A satisfies $|\lambda| < 1$.[37]

Now, using that A is symmetric, there is a basis in which A is diagonal:

$$A \sim \mathrm{diag}(1, \lambda_2, \dots, \lambda_N),$$

and to ask how quickly $A^n P$ approaches U, for any P, is "morally" the same as asking how small are the λ_i. (This can be made quantative and precise.) We continue the analysis therefore by looking further at this question.

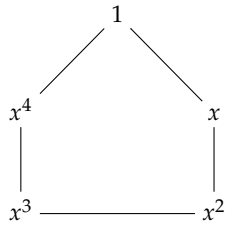
First, note that $A_{\rho_1 \oplus \rho_2} = A_{\rho_1} \oplus A_{\rho_2}$. As the regular representation decomposes as a direct sum of irreducible representations, we conclude the following: *the eigenvalues $\lambda_2, \dots, \lambda_N$ each an eigenvalue of some A_ρ, for some irreducible non-trivial representation ρ of G. Conversely, an eigenvalue of A_ρ for an irreducible representation of G is an eigenvalue of A.*

This gives, in principle, a way to analyze very closely the Cayley graph $\Gamma(G, S)$ if one has complete control over the irreducible representations. But, still, in practice, it could be a very difficult problem computationally. In certain cases, a great simplification occurs. Suppose that the set S is a union of conjugacy classes of G. For example, taking $G = S_n$ and S to be all transpositions and all 3-cycles gives a set satisfying all our assumptions. In this case, for any irreducible representation (ρ, V) with character χ we have $A_\rho \in \mathrm{End}_G(V)$ and, so, is a scalar, by Schur's lemma. Moreover, that scalar is the only eigenvalue of A_ρ and is given by the formula,

$$\frac{1}{\chi(1)} \mathrm{Tr}(A_\rho) = \frac{1}{d\chi(1)} \sum_{s \in S} \chi(s).$$

Note that this requires only knowing the character χ and avoids computation with matrices.

Example 22.8.1. For N odd and $S = \{1, -1\}$ the Cayley graph $\Gamma(\mathbb{Z}/N\mathbb{Z}, \{\pm 1\})$ is a circle. This is the case we considered previously for the random walk given by $x_{k+1} = x_k + b_k$ with $b_k = \pm 1$ with equal probability. (Note that the process $x_{k+1} = 3x_k + b_k$ is not equivalent to a Cayley graph. It is rather an object called a **Schreier graph** , and there is a theory for those as well.)



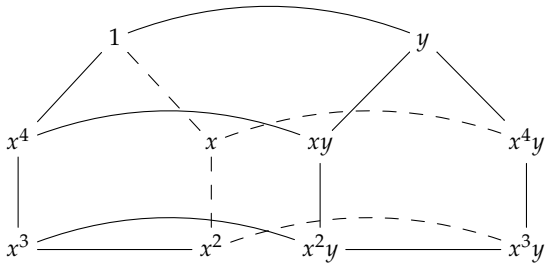
The Cayley graph $\Gamma(\mathbb{Z}/5\mathbb{Z}, \{\pm 1\})$

All the assumptions we made are satisfied and the eigenvalues A_ρ for the non-trivial characters are

$$\cos(2\pi j/N), \quad j = 1, \dots, N-1.$$

For the dihedral group $D_n = \langle x, y | y^2 = x^n = 1 \rangle$ we may take $S = \{x, x^{-1}, y\}$ and the Cayley graph $\Gamma(D_n, S)$ is the following:

[37] ≤ 1 is elementary; not equal to -1 is not hard and uses the fact that the graph is not bipartite.



The Cayley graph $\Gamma(D_5, \{x, x^{-1}, y\})$

However, in this case, the set S is not closed under conjugation so more work is required to find the eigenvalues. We assume that n is odd and we use the results about representations of dihedral groups in § 13.3. The non-trivial 1-dimensional representation receives the values $\psi(x^{\pm 1}) = 1, \psi(y) = -1$ and so gives the eigenvalue

$$\frac{1}{3}.$$

The irreducible 2-dimensional representations $\rho_a := \mathrm{Ind}_H^{D_n} \chi_a, a = 1, \ldots, (n-1)/2$, are determined by their values on the generators

$$x \overset{\rho_a}{\mapsto} \begin{pmatrix} \zeta^a & 0 \\ 0 & \zeta^{-a} \end{pmatrix}, \qquad y \overset{\rho_a}{\mapsto} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

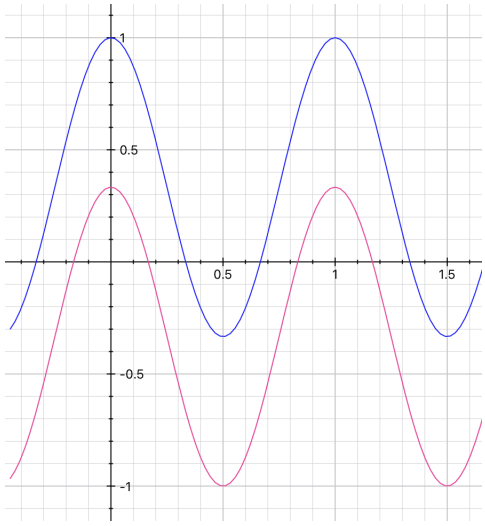
The corresponding operator A_{ρ_a} is thus given by the matrix

$$\frac{1}{3} \begin{pmatrix} 2\cos(2\pi a/n) & 1 \\ 1 & 2\cos(2\pi a/n) \end{pmatrix}.$$

Let $t = 2\cos(2\pi a/n)$. The characteristic polynomial of $3A_{\rho_a}$ equal to $x^2 - 2tx + (t^2 - 1)$ and has roots $t \pm 1$. Thus, the eigenvalues of the A_{ρ_a} are

$$\frac{1}{3}(2\cos(2\pi a/n) \pm 1), \quad a = 1, \ldots, (n-1)/2.$$

The graph of the functions $y = \frac{1}{3}(2\cos(2\pi x) \pm 1)$ looks as follows:



The fact that there are eigenvalues that are so close to 1 (e.g., $\frac{1}{3}(2\cos(2\pi x) + 1)$) is again responsible for the slow convergence to the uniform distribution, which is at the same rate as for the walk on the graph $\Gamma(\mathbb{Z}/n\mathbb{Z}, \{\pm 1\})$ that we analysed before.

22.9. **Riffle shuffles.** This is a famous problem that one can attack by similar techniques. The actual estimates are rather difficult and require full and detailed knowledge of the representation theory of the symmetric group. It is interesting, nonetheless, to see how the problem is set up and the first steps of the analysis.

A deck of cards, consisting of N cards ($N = 52$ in a usual deck) is split into two piles, one with k cards and the other with $N-k$ cards, with probability $\frac{1}{2^N}\binom{N}{k}$. Say, the left pile and the right pile. Then the cards from the two piles are interleaved randomly, where a card is chosen from the left pile with probability k/N and from the right pile with probability $(N-k)/N$. In the new pile the cards appear in a new order that is a permutation $\pi \in S_N$. Such a permutation is called, naturally enough, a **shuffle**, and the process of shuffling cards this way is called **riffle shuffle** or **dovetail shuffle**. After the shuffle, for some k, the cards $1,\dots,k$ appear in the deck in that order, but interleaved with the cards $k+1,\dots,N$ (that also appear in the deck in that order) in a random way. An example is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \dots & \dots & \dots & N-2 & N-1 & N \\ k+1 & 1 & 2 & k+2 & 3 & k+4 & \dots & k-2 & \dots & k-1 & N & k \end{pmatrix}$$

Experiments show that this is a good model for real-life card shuffles.

After n shuffles we get a certain probability distribution on S_N. If P is the original distribution, the distribution after n shuffles is P^{*n}. It is easy to understand the distribution P. We have $P(\pi) = 0$ if π is not a k-shuffle for any k, and $P(\pi) = 2^{-N}$ if π is a k-shuffle. But it is complicated to describe P^{*n} (and you can convince yourself of that by considering the case $n = 2$); more sophisticated methods are needed.

Similarly to the case of random walks on $\mathbb{Z}/p\mathbb{Z}$, routine arguments with Markov chains show that $P^{*n} \to U$ relative to the total variation norm. The question is how fast? Once more the main idea is to use the Diaconis-Shahshahani Lemma to get an estimate of the form

$$\|P^{*n} - U\|^2_{max} \leq \frac{1}{4} \sum_{\rho \neq \mathbb{1},\ \text{irred.}} \dim(\rho) \cdot \mathrm{Tr}((\hat{P}(\rho)^*)^n \cdot (\hat{P}(\rho))^n),$$

where now ρ runs over all irreducible representations of S_n.

The following table (their Q is our P) is taken from a paper of Bayer and Diaconis. It shows that 7 shuffles suffice to shuffle reasonably-well a deck of 52 cards and that, on the other hand, even 5 shuffles will exhibit significant bias towards particular permutations. .

Total variation distance for m shuffles of 52 cards

m	1	2	3	4	5	6	7	8	9	10
$\|Q^m - U\|$	1.000	1.000	1.000	1.000	0.924	0.614	0.334	0.167	0.085	0.043

22.10. **Rubik's cube.** Rubik's cube, or the Hungarian cube, is a well-known game. One is given a "scrambled" cube and using certain moves, one tries to unscramble the cube, bringing it to a position in which every face has the same colour.

To introduce group theory into the study of this game, one notes that there is a group G acting on the cube. The group G is generated by 6 basic moves a, b, c, d, e, f (each is a rotation of a certain "third of the

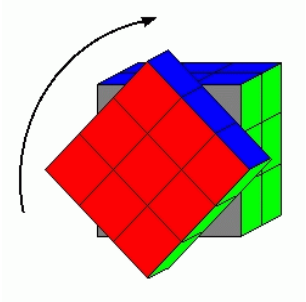


Figure 1. The Rubik Cube.

cube") and could be thought of as a subgroup of the symmetric group on $54 = 9 \times 6$ letters. It is called the **cube group**. The structure of this group is known. It is isomorphic to

$$(\mathbb{Z}/3\mathbb{Z}^7 \times \mathbb{Z}/2\mathbb{Z}^{11}) \rtimes ((A_8 \times A_{12}) \rtimes \mathbb{Z}/2\mathbb{Z}).$$

The order of the cube group is

$$2^{27} \cdot 3^{14} \cdot 5^3 \cdot 7^2 \cdot 11 = 43,252,003,274,489,856,000,$$

while the order of S_{54} is

$$2308436973392413804720927426830275810832785645718079411322880000000000000.$$

One is usually interested in solving the cube. Namely, reverting it to its original position. Since the current position was gotten by applying an element τ of G, in group theoretic terms we attempt to find an algorithm of writing every element in G in terms of the generators a, b, c, d, e, f since then also τ^{-1} will have such an expression, which is nothing else than a series of moves that returns the cube to its original position. It is natural do deal with the set of generators $a^{\pm 1}, b^{\pm 1}, \ldots, f^{\pm 1}$ (after all, why do 3 times a when you can do a^{-1}). A common question is what is the maximal number of basic operations that may be required to return a cube to its original position. Otherwise said, what is the diameter of the Cayley graph of G relative to the generators $\{a^{\pm 1}, b^{\pm 1}, c^{\pm 1}, d^{\pm 1}, e^{\pm 1}, f^{\pm 1}\}$? But more than that, is there a simple algorithm of finding for every element of G an expression in terms of the generators? The speed at which some people are able so solve the cube certainly suggests that the answer is yes! The current world record (June 2020) is 3.47 seconds, achieved by Yusheng Du from China in 2018.

The cube group is a rather complicated subgroup of S_{54}. For example, it has an element of order $1260 = 2^2 \cdot 3^2 \cdot 5 \cdot 7$. Usually, one denotes the moves not as we did, but by the letters u, d, l, r, f, b for *up, down, left, right, front, back*. The letter u signifies then rotating the upper face 90^0 clockwise if one looks straight at the face. Similarly, r means rotating the right face 90^0 clockwise if one looks straight at the face. In this notation, the element of order 1260 is $ru^2d^{-1}bd^{-1}$. Note that if we enumerate the 54 faces and performed this element we could encode it as a permutation and by decomposing it into a product of disjoint cycles easily check its order.

We introduced the notation u, d, f, b, l, r and the Cayley graph of the cube group relative to the generators $u^i, d^i, f^i, b^i, l^i, r^i$, $i = 1, 2, 3$. There is a rational for using this redundant set of generators; in practice, the moves $u^2, u^3 = u^{-1}$, for example, take almost the same time as u.

In cube solving competitions, cube scramblers are used. These are computer programs that produce a position of the cube and a set of instructions of how to get to it that judges use to create the cube positions to be solved. Naturally, we wish to have all cube positions given to the participants "equally hard", and also "hard enough" so that undeserving achievements will not be recorded as world-records. One needs to find a method that produces such positions. The scramblers are choosing randomly generators to provide directions

for creating the cube positions. However, we would like to guarantee that (with high probability) such sets of directions lead to equally hard positions that are also among the hardest possible.

The question of which position requires the most moves to solve was open for a long time and was finally settled by Rokicki et al. that determined this number to be 20. (This number is known as "God's number"; personally, I don't like this terminology.) The following table is taken from a paper of Rokicki; the first column indicates the minimal number of moves required to solve a position and the last column indicates the number of cube positions requiring this number. We ignore the middle column; it relates to the method of analysis used in their paper.

d	Canonical sequences	Positions
0	1	1
1	18	18
2	243	243
3	3,240	3,240
4	43,254	43,239
5	577,368	574,908
6	7,706,988	7,618,438
7	102,876,480	100,803,036
8	1,373,243,544	1,332,343,288
9	18,330,699,168	17,596,479,795
10	244,686,773,808	232,248,063,316
11	3,266,193,870,720	3,063,288,809,012
12	43,598,688,377,184	40,374,425,656,248
13	581,975,750,199,168	531,653,418,284,628
14	7,768,485,393,179,328	6,989,320,578,825,358
15	103,697,388,221,736,960	91,365,146,187,124,313
16	1,384,201,395,738,071,424	≈1,100,000,000,000,000,000
17	18,476,969,736,848,122,368	≈12,000,000,000,000,000,000
18	246,639,261,965,462,754,048	≈29,000,000,000,000,000,000
19	3,292,256,598,848,819,251,200	≈1,500,000,000,000,000,000
20	43,946,585,901,564,160,587,264	≈300,000,000

We see that the bulk of the cube positions require 18 moves. It is thus natural to perform the random process P and hope that P^{*n} is very closed to a distribution Q that has values, say, $Q(17) \approx Q(19) \approx 0.05$, $Q(18) \approx 0.90$ and otherwise $Q(i) \approx 0$. But, is it possible?? More precisely, what is

$$\min_n \|P^{*n} - Q\|_{max}.$$

I don't know the answer to that. (A careful analysis might require understanding the representations of the Cube group.) In real-life, the *Tnoodle scrambler program* is used by the World Cube Association to generate positions and the quality bar seems pretty low. At some point in time, they were OK with producing cube positions only guaranteed to require 11 moves or more, which seems rather bad. By simply running the program for say 1,000 times for each $n = 15 - 25$ and using fast cube-solvers, one could get a very reliable statistics on this question. The whole project shouldn't take more than a week to run a desktop computer.

23. Applications of group representations

This section provides some pointers to the literature. I will leave it to you to consult these references if you are interested. First, there are the two survey articles by T. Y. Lam, *"Representations of Finite Groups: A Hundred Years, Part I, and Part II"*. You can find the articles here:

http://www.ams.org/notices/199803/lam.pdf

http://www.ams.org/notices/199804/lam2.pdf

Secondly, there is the following post on Math overflow about "Fun applications of representations of finite groups", from which I have learned a lot myself.

https://mathoverflow.net/questions/11784/fun-applications-of-representations-of-finite-groups

I don't know if I would have used the adjective "fun", but there are certainly diverse and interesting applications. You would note, in particular, applications to:

(1) *Chemistry and Physics*, specifically quantum chemistry and quantum physics. For example, one user mentions "The symmetry group of a molecule controls its vibrational spectrum, as observed by IR spectrosocopy. When Kroto et al. discovered C60, they used this method to demonstrate its icosahedral symmetry." They suggest *Group Theory and Chemistry* by David M. Bishop as a reference. Another post suggests the book *Group Theory and Physics* by S. Sternberg for the connections to Physics quoting Sternberg saying that "molecular spectroscopy is an application of Schur's lemma". Another very convincing book is *Group theory and its applications to physical problems* by M. Hamermesh.

(2) *Combinatorics*. A lot of this is done through representations of the symmetric group and related groups. This is a topic to which many books, book chapters, and articles are devoted. The symmetric group plays a crucial role in combinatorics, of course. Mathscinet returns 455 references for searching for "Representation" and "symmetric group" in title, among which 14 are books.

(3) *Probability and Statistics*. Most notably, the book *Group representations in probability and statistics* by P. Diaconis that we have been following in these notes.

(4) Within *algebra*, the celebrated Feit-Thompson theorem that states that any finite simple non-abelian group has even order (equivalently, that any group of odd order is solvable), uses the following theorem of Frobenius, to which the only known proofs use representation theory.

A finite group G is called a **Frobenius group** with Frobenius kernel K and Frobenius complement H if G has a subgroup H, such that for any $g \notin H$ we have $H \cap gHg^{-1} = \{1\}$. One lets in this case

$$K = \{1\} \cup (G - \bigcup_{g \in G} gHg^{-1}).$$

K is called the **Frobenius kernel**. If we let G act by left multiplication on the left cosets G/H, the stabilizer of gH is gHg^{-1} and the condition on H can be interpreted as saying that the only element in G that has more than 1 fixed point is the identity. The subgroup K consists of the identity and the elements that have no fixed points.

An example of a Frobenius group is the group of invertible affine linear transformations of the line $\{ax + b\}$ with H being the linear transformations $\{ax\}$ (the stabilizer of the point 0) and K the translation maps $x \mapsto x + b$. If the line is over a finite field $\mathbb{F}$, we can also write this group as $\{\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right) : a \in \mathbb{F}^\times, b \in \mathbb{F}\} \cong \mathbb{F} \rtimes \mathbb{F}^\times$.

Theorem 1 (Frobenius' theorem) *Let G be a Frobenius group with Frobenius complement H and Frobenius kernel K. Then K is a normal subgroup of G, and G is the semidirect product $K \rtimes H$.*

The hard part is to show that K is a group! As we have indicated above, we can reformulate the theorem.

Theorem 2 (Frobenius' theorem, equivalent version) *Let G be a group of permutations acting transitively on a finite set X, with the property that any non-identity permutation in G fixes at most one point in X. Then the set of permutations in G that fix no points in X, together with the identity, is closed under composition.*

Apparently, there is still no proof of these theorems that avoids using group representations in an essential way. Although, recently, Terrence Tao gave a proof that only uses character theory for finite groups. I have learned much about this from reading Tao's blog

https://terrytao.wordpress.com/2013/04/12/the-theorems-of-frobenius-and-suzuki-on-finite-groups/

Finally, but still within the realm of pure algebra, group representations have a lot to do with the study of simple groups. The classification of simple groups puts them in large families ($\mathbb{Z}/p\mathbb{Z}, A_n, \mathrm{PSL}_n(\mathbb{F}), \dots$) but some escape this classification and fall into a category to themselves: the sporadic simple groups. There are finitely many such groups (27, in fact). The largest simple group is the Monster group, whose order is

$$808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000.$$

Its existence is a non-trivial fact. Before constructing the Monster, mathematicians suspected its existence and in fact predicted the dimensions of some of its smallest irreducible representations as $1,196883$ and 21296876, and were able, more generally, to work out its character table. John McKay (1939-2022) made the audacious observation that those numbers are related to Fourier coefficients of the j-function, a function appearing in the theory of elliptic curves, which is part of number theory. Following that, precise conjectures were made by Conway and Norton, going under the name of "Moonshine".

Some of the key aspects of these conjectures were proven by R. Borcherds, a work that got him the Fields prize in 1998.

(5) In number theory, representations of groups play a central role. The subject of automorphic forms is really about the representation theory of certain infinite groups. At a more accessible level, group representations play an important role in the study of L-functions, and in the study of equations over finite fields (for example, Gauss sums can be viewed as Fourier transforms). For a concrete example, in a different direction, we might mention Roth's theorem:

A subset A of the natural numbers is said to have positive **upper density** if

$$\limsup_{n\to\infty} \frac{|A\cap\{1,2,3,\dots,n\}|}{n} > 0.$$

Roth's theorem on arithmetic progressions states that a subset of the natural numbers with positive upper density contains a 3-term arithmetic progression; namely, there are positive integers x, d such that $x, x+d, x+2d$ belong to A. This is a surprisingly difficult theorem, and many people have worked on various generalizations of it. This is a subject of ongoing research. Roth's proof, as well as current day developments, use heavily Fourier analysis on the group $\mathbb{Z}/n\mathbb{Z}$.

Part 4. SEMISIMPLE RINGS AND MODULES

In this Part 4 of the notes, we discuss semisimple modules and rings in §§24-28. We will develop the theory systematically for left modules, but of course it can be developed in the same way for right modules. We will sometimes be imprecise in our terminology and refer to a "module" when we (always) mean a "left module". In §29 we discuss central simple algebras and the Brauer group. The Brauer group is connected to many deep phenomena in algebra, number theory and geometry and we barely scratch the subject.

24. Semisimple modules

24.1. **Simple modules and Schur's Lemma.** *In this section we define simple left modules and study their homomorphisms. We prove that the endomorphism ring of a simple module is a division algebra.*

Let R be a ring, not necessarily commutative. Let $M \in {}_R\mathbf{Mod}$. Recall that M is called **simple** (or **irreducible**) if $M \neq \{0\}$ and any submodule of M is either $\{0\}$ or M itself.

Lemma 24.1.1. *M is simple if and only if $M \cong R/I$ where I is a maximal left ideal of R.*

Proof. If M is simple, choose $m \in M, m \neq 0$. Then $N = Rm$ is a non-zero submodule of M, hence equal to M. The map

$$R \to M, \quad r \mapsto rm$$

is a map of R modules and its kernel I is a left R-module contained in R; that is, the kernel is a left ideal I and

$$R/I \cong M.$$

The left ideals $J \supset I$ correspond to submodules of M. Thus, such J is either I or R and it follows that I is maximal.

Conversely, if I is maximal, $M = R/I$ is an R-module. The same argument shows that if I is maximal, M has no non-trivial submodules, hence is simple. □

Example 24.1.2. Let D be a division ring. Using Morita's equivalence we see that the simple modules over $M_n(D)$ correspond to simple modules over D, of which there is a unique one. Namely, D itself. It follows that, up to isomorphism, the only simple module over $M_n(D)$ is D^n (column vectors) on which matrices act as usual. Further,

$$D^n \cong M_n(D)/\left\{ \begin{pmatrix} 0 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & & & \vdots \\ 0 & * & \cdots & * \end{pmatrix} \right\} = M_n(D)/I.$$

Lemma 24.1.3. (Schur) *Let M, N be simple R-modules, $f\colon M \to N$ a homomorphism of modules. Then, either $f = 0$ or f is an isomorphism.*

Proof. $\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$ are both submodules, hence either $\mathrm{Ker}(f) \neq 0$, which implies both $\mathrm{Ker}(f) = M$ and $f = 0$. Or, $\mathrm{Ker}(f) = 0$ and then $\mathrm{Im}(f) \neq 0$, which implies that $\mathrm{Im}(f) = N$ and hence that f is an isomorphism. □

Corollary 24.1.4. *Let M be a simple R-module, then $\mathrm{End}_R(M)$ is a division ring.*

Proof. Any $f \in \mathrm{End}_R(M), f \neq 0$, is an isomorphism, hence invertible, meaning has a two sided inverse. □

Example 24.1.5. Again by Morita's equivalence, we have

$$\mathrm{End}_{M_n(D)}(D^n) \cong \mathrm{End}_D(D,D) \cong D$$

(acting from the right); the latter isomorphism takes f to $f(1)$.

24.2. **Semisimple modules.** *In this section we define semisimple modules and find several characterizations of them. These allow us to conclude that the collection of semisimple modules is closed under direct sums, homomorphic images and submodules. We prove that every module is semisimple if and only if the ring itself is semisimple (viewed as a module over itself).*

A module $M \in {}_R\mathbf{Mod}$ is called **semisimple** if M is isomorphic to a direct sum of simple R-modules,

$$M \cong \oplus_{i\in I} M_i, \quad M_i \text{ simple, } \forall i \in I.$$

Here I is an index set that is allowed to be infinite. I is also allowed to be empty, hence the zero module is semisimple (though not simple).

Theorem 24.2.1. *The following are equivalent for an R-module M:*

1. *M is semisimple.*
2. *M is generated by its simple submodules.*
3. *For every submodule $N \subsetneqq M$ there exists a submodule $\{0\} \neq P \subseteq M$ such that $N \cap P = \{0\}$.*
4. *For every submodule $N \subseteq M$ there exists a submodule $P \subseteq M$ such that $M = N \oplus P$ (internal direct sum).*

Proof. We begin by showing $(1) \Rightarrow (2)$. If $M = \oplus_{i\in I} M_i$, M_i simple $\forall i \in I$, then M is generated by the simple submodules M_i (where M_i is identified with the vectors in $\oplus_{i\in I} M_i$ with non-zero coordinates only at the i-th place).

Next we show $(2) \Rightarrow (3)$. We assume M is generated by its simple submodules, say $\{M_i : i \in I\}$. Since $N \neq M$, there is some M_{i_0} such that $M_{i_0} \not\subset N$. Then, as $N \cap M_{i_0}$ is a proper submodule of M_{i_0} and M_{i_0} is simple, we must have $N \cap M_{i_0} = \{0\}$; we may take $P = M_{i_0}$.

Assume (3) and let N be a submodule of M and let P be as in (3). We may assume $N \neq M$. Let

$$\Sigma = \{Q \subset M \text{ submodule} : Q \cap N = \{0\}\}.$$

Then Σ is non-empty as $P \in \Sigma$ and Σ is partially ordered under inclusion. Every chain in Σ has a supremum which is the union of its members. Thus, by Zorn's Lemma, Σ has a maximal element P^+. Suppose that $N + P^+ \neq M$, then, by (3), there is some non-zero submodule P_1 of M such that $(N + P^+) \cap P_1 = \{0\}$. Then $P^+ + P_1 \supsetneqq P^+$ and $N \cap (P^+ + P_1) = \{0\}$. This contradicts the maximality of P^+. In summary, we proved that $N \cap P^+ = \{0\}, N + P^+ = M$, and we conclude

$$M = N \oplus P^+.$$

It remains to prove $(4) \Rightarrow (1)$. Pick some index set I such that $|I| > |M|$. Let

$$\Sigma = \{\oplus_{i\in I} H_i \subset M : H_i \text{ a simple submodule or } \{0\}\}.$$

For two such direct sums $\oplus_{i\in I} H_i, \oplus_{i\in I} H_i'$, we say that $\oplus_{i\in I} H_i \leq \oplus_{i\in I} H_i'$ if for all i, $H_i \subset H_i'$.(Note though that since H_i' is either 0 or simple, the only possibility is to make some of the H_i that are zero into non-zero simple submodules H_i' of M.) Then Σ is not empty (it contains the sum of zero modules) and satisfies the conditions of Zorn's lemma. There is thus some maximal element N of Σ, say

$$N = \oplus_{i\in I} H_i.$$

For cardinality reasons, there is some i_0 such that $H_{i_0} = \{0\}$.

Suppose $N \neq M$ and, using (4), let $P \subset M$ be a submodule such that

$$N \oplus P = M.$$

We are going to use P to construct a simple submodule that can be added to N as a direct sum and thus contradict its maximality. Let $a_1 \in P$, $a_1 \neq 0$, $P_1 = Ra_1 \cong R/I_1$ for some left ideal I_1 of R. Let $I \supset I_1$ be a maximal ideal and let P_2 be a submodule such that

$$(N \oplus P_1) \oplus P_2 = M.$$

Note that $(N \oplus P_1) \oplus P_2 \supseteq N \oplus Ia_1 \oplus P_2$ and

$$M/(N \oplus Ia_1 \oplus P_2) \cong R/I$$

is a simple R-module. Now, there exists an R submodule $P_3 \subset M$ such that

$$N \oplus Ia_1 \oplus P_2 \oplus P_3 = M,$$

and $P_3 \cong R/I$ is simple. We now find a bigger direct sum

$$N \oplus P_3$$

which we an write as $\oplus_{i \in I} H_i'$, with $H_i' = H_i$ for $i \neq i_0$ and $H_{i_0}' = P_3$. This contradicts the maximality of N. Therefore, $N = M$ and we find that M is a direct sum of simple modules. □

Corollary 24.2.2. *The following hold:*

(1) A direct sum of semisimple modules is semisimple.
(2) A homomorphic image of a semi-simple module is semisimple.
(3) A submodule of a semisimple module is semisimple.

Proof. (1) is already clear from the definition.

For (2), note that if M is generated by simple submodules M_i, and $f\colon M \to N$ is a surjective homomorphism, then $N = f(M)$ is generated by the modules $f(M_i)$, where each $f(M_i)$ is either zero or isomorphic to M_i.

To see (3), let N be a submodule of a semisimple module M. There is a submodule P such that $M = N \oplus P$ then $N \cong M/P$ is a homomorphic image of M, hence semisimple. □

Corollary 24.2.3. *Let R be a ring. Then every R-module is semisimple if and only if R is a* **semisimple ring**, *meaning R is semisimple as an R-module.*

Proof. One direction is obvious. Consider the converse: If R is a semisimple ring then $\oplus_{i \in I} R$ is a semisimple R-module, the sum over any index set I. But any R-module is a homomorphic image of such a sum: if M is an R-module M is a homomorphic image of $\oplus_{m \in M} R$, where an element $r \in R$ in the m-th coordinate is sent to rm under the homomorphism. Thus, any R-module is semisimple. □

Remark 24.2.4. One can prove that a ring R is a simple R-module if and only if R is a division ring. (The natural argument produces for $a \neq 0$ an element b such that $ba = 1$. One needs to show that then also $ab = 1$. Consider $abab$. In a general ring it is possible that for a given a there is a b such that $ba = 1$ but $ab \neq 1$.)

24.3. **Semisimple rings.** *Semisimple rings are very accessible because every module over them is semisimple. If R is a semisimple ring one can say a bit more about its structure and that is what we do in this section.*

Proposition 24.3.1. *Let R be a left semisimple ring then R is a direct sum of* finitely many *minimal left ideals.*

Proof. Firstly, a (left) R-submodule of R is a (left) ideal and it is simple if and only if that ideal is minimal, meaning not zero and not properly containing any other non-zero left ideal. Thus, if R is semisimple, we have an isomorphism of left R-modules,

$$R \cong \oplus_{i\in I} J_i,$$

where each J_i is a minimal left ideal. In particular, under this decomposition, $1 = \oplus e_i$, $e_i \in J_i$, where all e_i but finitely many are 0. Now, for any r we have

$$r = \oplus r_i, \qquad r = r\cdot 1 = r\cdot \oplus e_i = \oplus r e_i.$$

We conclude that $r_i = re_i$ and so if $e_i = 0$ also $r_i = 0$. It follows that the sum $\oplus_{i\in I} J_i$ is actually a finite sum and for every i, $e_i \neq 0$. □

For the next corollary, refer to § 25.1 for the definition of artinian modules and some results used in the proof.

Corollary 24.3.2. *Let R be a left semisimple ring then R is left artinian and left noetherian.*

Proof. We have $R = \oplus_{i=1}^N J_i$, where J_i are minimal left ideals. For any $1 \leq a \leq N-1$ we have an exact sequence of R modules

$$0 \to \oplus_{i=1}^a J_i \to \oplus_{i=1}^{a+1} J_i \to J_{a+1} \to 0.$$

Note that J_{a+1} has no non-trivial submodules hence is both artinian and noetherian. This shows by induction on a that each $\oplus_{i=1}^{a+1} J_i$ is artinian (resp., noetherian) and thus so is R. □

Example 24.3.3. If D is a division ring then its only simple module is D (because any simple module is D/I and there are no non-trivial left ideals). As D is a simple module over itself, it is semisimple. Thus, every module over D is a direct sum of simple modules, namely of copies of D. This is something we already used without ever proving it (just alluding to "linear algebra").

Using Morita equivalence we conclude that $M_n(D)$ is a semisimple ring. In fact, as a module it is isomorphic to the following direct sum of simple modules

$$D^n \oplus \cdots \oplus D^n \quad (n \text{ times}).$$

(The i-th column in a matrix being the i-th copy of D^n.)

One can prove that if $R_1, \ldots, R_n$ are semisimple rings then so is $R = R_1 \times \cdots \times R_n$. This is quite easy when we note that we can view any R_i-module as an R-module by letting R act through the projection on its i-th component R_i. A simple R_i-module is then a simple R-module. Moreover, if each R_i is artinian so is R.

Therefore, we may conclude that if D_i are division algebras and n_i are positive integers

$$M_{n_1}(D_1) \times \cdots \times M_{n_r}(D_r)$$

is a semisimple artinian ring. The Artin-Wedderburn Theorem 28.2.1 asserts that all non-zero semisimple artinian rings are like that!

Example 24.3.4. The simple $\mathbb{Z}$-modules are $\mathbb{Z}/p\mathbb{Z}$, p a prime. In particular $\mathbb{Z}$ is not a semisimple ring. You can extend this argument to many other examples of PIDs, for example $\mathbb{F}[x]$, where $\mathbb{F}$ is a field.

Example 24.3.5. Let G be a finite group. Maschke's theorem states that $\mathbb{C}[G]$ is a semisimple ring as it decomposes as a sum of irreducible representations (=simple $\mathbb{C}[G]$-modules) over itself. The theorem is true for $k[G]$, where k is any field such that $(\mathrm{char}(k), \sharp G) = 1$, although a different proof is needed.

Maschke's theorem may fail without the assumption on the characteristic of k. For example, $\mathbb{F}_2[G]$ is not semisimple if $G = \langle x \rangle$ is a cyclic group of order 2. Indeed, we have isomorphism of rings $\mathbb{F}_2[G] \cong \mathbb{F}_2[x]/(x^2-1) = \mathbb{F}_2[x]/(x-1)^2 \cong \mathbb{F}_2[t]/(t^2)$. The submodule, i.e. ideal, (t) does not have a direct sum complement.

25. The Jacobson radical

In this section we introduce the Jacobson radical of a ring and of a module and use it to study semisimplicity. Certain functorial properties of the Jacobson radical and different characterizations of it are provided. The Jacobson radical is also used in the following section (Nakayama's Lemma).

25.1. **Artinian rings and modules.** Let R be a a ring and $M \in {}_R\mathbf{Mod}$ a left R-module. We call M an **artinian module** if any descending chain of submodules of M stabilizes. Namely, if

$$M \supseteq M_0 \supseteq M_1 \supseteq M_2 \supseteq \dots$$

are modules then there exists an integer N such that

$$M_i = M_N, \quad \forall i \geq N.$$

We call R a left **artinian ring** if it is an artinian left R-module. Namely, R is a left artinian ring if any descending chain of left ideals of R stabilizes.

There is an analogous definition of artinian right modules and so a definition of a right artinian ring. We remark that a ring may be left artinian and not right artinian (or the other way around).

Example 25.1.1. Here are some examples.

(1) Every ring with finitely many elements is artinian.

(2) Let D be a division algebra, for example a field. Let R be a ring containing D and of finite dimension as a left module over D. Then R is an artinian ring. Indeed, any ideal I of R is in particular a left module over D and thus has a dimension $\dim_D(I)$, which is finite and non-negative. If $I \supsetneq J$ are left ideals then $I = J$ if and only if $\dim_D(I) = \dim_D(J)$. It follows that there cannot be an infinite chain of strictly descending ideals.

In particular, $M_n(D)$ is an artinian ring for every $n \in \mathbb{N}, n \geq 1$. This is a very typical way in which artinian rings arise.

(3) If R is a PID and $f \in R$ a non-zero element, then $R/\langle f \rangle$ is an artinian R-module and an artinian ring. The sub R-modules are ideals of $R/\langle f \rangle$ and those correspond to the finitely many divisors of f, up to the relation of being associate.

This, too, is a very typical way in which artinian rings arise. A very simple example is $\mathbb{C}[x]/(x^n)$.

(4) $\mathbb{Z}$ is a noetherian ring, but is not an artinian ring. For example, we have the strict inclusions

$$\mathbb{Z} \supsetneq 2\mathbb{Z} \supsetneq 4\mathbb{Z} \supsetneq 8\mathbb{Z} \supsetneq \dots.$$

It is natural to ask if, conversely, there are artinian rings that are not noetherian. The next theorem answers that.

Theorem 25.1.2. (Hopkins-Levitzki) *Let R be a left (right) artinian ring then R is left (resp., right) noetherian ring.*

This is not true for modules as the next example shows.

(5) Let p be a prime number. $\cup_{k=1}^{\infty}\frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$ is an artinian $\mathbb{Z}$-module, but is not a noetherian $\mathbb{Z}$-module.

(6) $\mathbb{Q}$ is neither artinian nor noetherian as a $\mathbb{Z}$-module. For example, the following is an infinite chain of $\mathbb{Z}$-submodules extending in both directions:

$$\cdots \supsetneq \frac{1}{4}\mathbb{Z} \supsetneq \frac{1}{2}\mathbb{Z} \supsetneq \mathbb{Z} \supsetneq 2\mathbb{Z} \supsetneq 4\mathbb{Z} \supsetneq 8\mathbb{Z} \supsetneq \cdots$$

Lemma 25.1.3. *Let M be an artinian module and let $\{M_i : i \in I\}$ be a collection of submodules of M such that $\cap_{i\in I}M_i = \{0\}$. Then there exists finitely many elements $i_1,\dots,i_r$ in I such that $\cap_{j=1}^{r}M_{i_j} = \{0\}$.*

Proof. Pick any $i_1 \in I$. If $M_{i_1} = \{0\}$ we are done. Else, pick some i_2 such that $M_{i_1} \cap M_{i_2} \subsetneq M_{i_1}$. Such must exists, else $\cap_{i\in I}M_i = M_{i_1} \neq \{0\}$. If $M_{i_1} \cap M_{i_2} = \{0\}$ we are done. Else, find an M_{i_3} such that $M_{i_1} \cap M_{i_2} \cap M_{i_3} \subsetneq M_{i_1} \cap M_{i_2}$. And so on. If this process doesn't stop we get an infinite chain of strictly descending submodules of M, contradicting the artinian property. If the process stops then we have arrived at a choice $i_1,\dots,i_r$ such that $\cap_{j=1}^{r}M_{i_j} = \{0\}$. □

Proposition 25.1.4. (Artinian induction) *Let M be an artinian R-module and $\mathscr{P}$ a property of R-modules. If there is a non-zero submodule of M with property $\mathscr{P}$ then there is a non-zero submodule of M that is minimal relative to having this property. In particular, this applies to artinian rings R and properties of their ideals.*

Proof. If not, then one can construct inductively a strictly descending sequence of submodules of M. □

Exactly in the same way we have

Proposition 25.1.5. (Noetherian induction) *Let M be a noetherian R-module and $\mathscr{P}$ a property of R-modules. If there is a submodule of M, properly contained in M, with property $\mathscr{P}$ then there is a submodule of M, property contained in M, that is maximal relative to having this property. In particular, this applies to noetherian rings R and properties of their ideals.*

Lemma 25.1.6. *Let $0 \longrightarrow M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \longrightarrow 0$ be an exact sequence of R-modules. Then M_2 is artinian if and only if both M_1 and M_3 are artinian.*

Proof. The proof is very similar to the proof of Lemma 7.1.5. □

Corollary 25.1.7. *If R is an artinian ring then any finitely generated R-module is an artinian R-module.*

Proof. First argue for R^n by induction on n, using the lemma. And then, apply the lemma again to prove it for quotients of R^n for some n, namely for finitely generated R-modules. □

Corollary 25.1.8. *If $\{M_i\}_{i=1}^{n}$ are simple R-modules then $\oplus_{i=1}^{n}M_i$ is an artinian R-module.*

Proof. Argue by induction on n, using the lemma and noting that every M_i is artinian. □

25.2. **The Jacobson radical of a module.** Let $M \in {}_R\mathbf{Mod}$. We define the **Jacobson radical** of M to be the submodule

$$J(M) = \cap_{N \text{ max.}} N,$$

where the intersection is over all maximal submodules N of M. A submodule N is called **maximal** if there is no submodule N' such that $N \subsetneqq N' \subsetneqq M$; equivalently, if M/N is a simple module. If there are no maximal submodules then this is an intersection over an empty index set and therefore it is, by definition, equal to M.

An example of particular importance is when $M = R$. In that case

$$J(R) = \cap_{I \lhd R \text{ max. ideal}} I,$$

is an ideal of R – the intersection of all the maximal left ideals.[38]

Proposition 25.2.1. *Let M be a semisimple R-module then $J(M) = \{0\}$.*

Proof. Since M is semisimple, $M = \oplus_{i \in I} M_i$, a sum of simple R-modules. If $I = \emptyset$ then $M = 0 = J(M)$. if I has cardinality 1 then M is simple and its only maximal submodule is $\{0\}$ and thus $J(M) = \{0\}$. Assume that $|I| > 1$. (The following works of $|I| = 1$, but might be a bit confusing in that case.)

For every $i \in I$, $M^i := \oplus_{j \neq i} M_j$ is a submodule and, since $M/M^i \cong M_i$, it is a maximal submodule. Now, $\cap_{i \in I} M^i = \{0\}$ so also $J(M) = \{0\}$. □

We have the following proposition that will be instrumental later on.

Proposition 25.2.2. *Let M be an artinian R-module. If $J(M) = \{0\}$ then M is semisimple and, in fact, M is a finite direct sum of simple R-modules.*

Note the particular case when $M = R$. We find that if R is an artinian ring and $J(R) = \{0\}$ then R is a semisimple ring and, as we have already seen, a finite direct sum of minimal left ideals.

Proof. The case $M = \{0\}$ is trivial. Suppose $M \neq \{0\}$. Since $J(M) = \{0\}$, there exists a maximal submodule of M and thus, since R is artinian, by Lemma 25.1.3 there exists finitely many maximal submodules $M_i, i = 1, \ldots, n$ $(n \geq 1)$, such that $\cap_{i=1}^n M_i = \{0\}$. We consider the inclusion

$$M \hookrightarrow M/M_1 \oplus \cdots \oplus M/M_n.$$

Each M/M_i is a simple R-module, thus the sum on the right hand side is a semisimple module and, consequently, by Corollary 24.2.2, so is M. Therefore,

$$M = \oplus_{i \in I} H_i,$$

a direct sum of simple R modules. Choose a linear order of I, $I = \{i_0 < i_1 < i_2 < \cdots\}$ (I need not be countable, though). Let $N_n = \oplus_{i \geq i_n} H_i$. Then N_n is a strictly decreasing sequence of submodules of M that therefore must stabilize. It follows that I is finite. □

Remark 25.2.3. Note that there is a problem in extending the argument to prove that if $J(M) = \{0\}$ then M is semisimple. Indeed, we cannot say that $M \hookrightarrow \oplus_i M/M_i$ where M_i are the maximal submodules of M, because we dont know that the image of m has only finitely many non-zero coordinates (if it always does then the argument goes through to show that M is semisimple). This is not a technical issue, but a true obstacle: take the ring $\mathbb{Z}$, whose Jacobson radical is $\{0\}$; it is not semisimple.

Proposition 25.2.4. *The Jacobson radical has the following properties:*

(1) Let $f\colon M \to N$ be a module homomorphism then $f(J(M)) \subseteq J(N)$.
(2) $J(M/J(M)) = \{0\}$.
(3) $J(R)$ is a two-sided ideal of R.

[38] If $R \neq \{0\}$ then there always is a maximal left ideal. Compare it for the situation for module, for example to the case of $\mathbb{Q}$ that doesn't have maximal $\mathbb{Z}$-submodules. Why are these two cases different?

(4) $x \in J(R)$ if and only if for all $r \in R$, $\exists u \in R, u(1-rx) = 1$.

Proof. By the first isomorphism theorem for modules, very homomorphism is a composition of a surjection followed by an injection, and so it is enough to prove the first statement when f is surjective, or injective.

Suppose first that f is surjective. Let $x \in J(M)$ and let $N' \subset N$ be a maximal submodule (if none exists then $J(N) = N$ and there is nothing to prove). As f is surjective, $f^{-1}(N')$ is a maximal submodule of M and so $x \in f^{-1}(N')$, which implies $f(x) \in N'$. We proved $f(J(M)) \subseteq J(N)$.

Assume now that f is injective and so, without loss of generality, $M \subseteq N$. Let $x \in J(M)$ and N' a maximal submodule of N. If $N' \supseteq M$ then $x \in N'$. In the other case, $N'' := N' \cap M$ is a proper submodule of M, and $M/N'' \hookrightarrow N/N'$. As N/N' is a simple module, we must have that $M/N'' \cong N/N'$ and so M/N'' is simple. That is, N'' is a maximal submodule of M in this case and again $x \in N'$. We proved that $f(J(M)) \subseteq J(N)$.

The second statement is easy. Use the correspondence theorem. The maximal submodules of $M/J(M)$ are in bijection with maximal submodules of M (because they all contain $J(M)$). In fact, let $\pi\colon M \to M/J(M)$ be the canonical map. Then,

$$\cap_{N<M/J(M) \text{ max.}} N = \pi(\cap_{N<M \text{ max.}} N) = \pi(J(M)) = \{0\}.$$

To get the third statement, let $a \in R$ and consider the homomorphism of left R-modules

$$f\colon R \to R, \quad r \mapsto ra.$$

We have $f(J(R)) \subseteq J(R)$ but that just means that $J(R)a \subseteq J(R)$ for any $a \in R$. Namely, $J(R)$ is a two-sided ideal.

Now for the last statement. Let $x \in J(R)$ and suppose that for some $r \in R$, $1-rx$ does not have a left inverse. Then $R(1-rx)$ is a proper sub R-module of R, i.e. a proper left ideal. Thus, there is a maximal left ideal I such that $1-rx \in I$. However, $x \in I$ and it follows that $1 \in I$. Contradiction.

Conversely, suppose that $1-rx$ always has a left inverse. Let I be a maximal left ideal. If $x \notin I$ then the left R-module $I + Rx = R$. This implies that for some $a \in I, r \in R, a + rx = 1$. Then $a = 1 - rx$ is in I and left invertible, implying $I = R$. Contradiction. □

Note that we can restate (4) and say that

$$J(R) = \{z \in R : 1 - rz \text{ has a left inverse}, \forall r \in R\}.$$

It is not hard to check that if we define a right ideal $\tilde{J}(R)$ as the intersection of all maximal right ideals of R then we would again find that it is a two-sided ideal, now characterized by

$$\tilde{J}(R) = \{z \in R : 1 - zr \text{ has a right inverse}, \forall r \in R\}.$$

Now note that if an element a in a ring R has both a left inverse, say $ba = 1$, and a right inverse, say $ac = 1$, then $b = c$, because

$$b = b(ac) = (ba)c = c.$$

Moreover, b is uniquely determined: If $b'a = 1$ then

$$b' = b'(ab) = (b'a)b = b.$$

Thus, in the situation where a has both a left and a right inverse they must be the same and we denote them a^{-1}.

Let $z \in J(R), r \in R$. Then $zr \in J(R)$, because $J(R)$ is a two-sided ideal and so there is a $u \in R$ such that $u(1-zr) = 1$ and so u has a right inverse. But, also $u = 1 + uzr = 1 - (-uzr)$ and $-uzr \in J(R)$ so u has a left inverse too. It follows that u is a unit and we can write $(1-zr) = u^{-1}$ and conclude that $(1-zr)u = 1$. It follows that $z \in \tilde{J}(R)$. And conversely. We deduce:

Corollary 25.2.5. *$J(R)$ is also equal to the intersection of all maximal right ideals of R.*

Remark 25.2.6. One can prove the following characterization of $J(R)$, which we leave as an exercise. Any element of $J(R)$ acts as zero on any semi-simple R-module and this property characterizes $J(R)$. Applying this observation to the group ring $\mathbb{C}[G]$ of a finite group G and its regular representation (which is a faithful $\mathbb{C}[G]$-module), one conclude that $J(\mathbb{C}[G]) = \{0\}$.

26. Nakayama's Lemma and applications

26.1. **Nakayama's lemma.**

Lemma 26.1.1. *Let $N \in {}_{\mathbf{R}}\mathbf{Mod}$ be a finitely generated R-module. Then*

$$J(R)N = N \implies N = 0.$$

Proof. Let n be the minimal integer such that

$$N = Ra_1 + \cdots + Ra_n$$

for some $a_1, \ldots, a_n \in N$. Suppose that $N \neq \{0\}$, so $n \geq 1$. As $N = J(R)N$ and $J(R)$ is also a right ideal, we have, $a_1 \in N = J(R)N = J(R)a_1 + \cdots + J(R)a_n$, and that means that we can write

$$a_1 = j_1 a_1 + \cdots + j_n a_n, \quad \text{for some } j_i \in J(R).$$

Using that $(1 - j_1)$ is invertible, we find that

$$a_1 = (1 - j_1)^{-1} j_2 a_2 + \cdots + (1 - j_1)^{-1} j_n a_n,$$

and thus

$$N = Ra_2 + \cdots + Ra_n.$$

This is a contradiction and so it must be that $N = 0$. □

26.2. **Applications of Nakayama's Lemma.**

Proposition 26.2.1. *Let R be a ring and $x_1, \ldots, x_n$ elements of a finitely generated R-module M. Then*

$$x_1, \ldots, x_n \text{ generate } M \text{ over } R \iff \bar{x}_1, \ldots, \bar{x}_n \text{ generate } M/J(R)M \text{ over } R.$$

Proof. The implication $\Rightarrow$ is clear. In the other direction, let N be the module generated by $x_1, \ldots, x_n$ in M. Then, since

$$\frac{N + J(R)M}{J(R)M} = \frac{M}{J(R)M},$$

by the correspondence theorem for $M \to M/J(R)M$, we may conclude that $N + J(R)M = M$ and we find that R-module M/N satisfies

$$J(R)(M/N) = (J(R)M + N)/N = M/N.$$

This implies $M/N = \{0\}$ and thus that $M = N$. □

Corollary 26.2.2. *Let R be a commutative local ring with a maximal ideal $\mathfrak{m}$. Let N be a finitely generated R-module and $x_1, \ldots, x_n$ elements of N that generate $N/\mathfrak{m}N$. Then $x_1, \ldots, x_n$ generate N.*

Proof. This is an immediate conclusion from the proposition because in this case $\mathfrak{m} = J(R)$. □

Remark 26.2.3. In Theorem 31.2.1, we will show using Nakayama's lemma that any projective finitely-generated module over a commutative local ring is free.

Theorem 26.2.4. *Let R be a left artinian ring. Then $J(R)$ is the largest left ideal of R that is* **nilpotent**. *That is, for some $n \geq 1$ we have $J(R)^n = \{0\}$.*[39]

Proof. Let I be an ideal such that $I^n = \{0\}$ and let $x \in I$. The argument "$x^n = 0$ implies x belongs to any maximal ideal because any maximal ideal is prime", doesn't work in the non-commutative case. (For a start, the quotient by a left ideal has no ring structure in general.) We have to argue differently.

For any $r \in R, rx \in I$ and so $(rx)^n = 0$. We then find

$$(1 + rx + \cdots + (rx)^{n-1})(1 - rx) = 1 - (rx)^n = 1,$$

and therefore $(1 - rx)$ has a left inverse. It follows from Proposition 25.2.4 that $x \in J(R)$. Therefore, $I \subseteq J(R)$.

To show that $J(R)$ itself is nilpotent, consider the sequence of left ideals

$$J(R) \supseteq J(R)^2 \supseteq J(R)^3 \supseteq \dots$$

Using the artinian property, we have for some n,

$$J(R)J(R)^n = J(R)^{n+1} = J(R)^n.$$

If we knew that $J(R)$ is finitely generated, we could have deduced that $J(R)^n$ is finitely generated and apply Nakayama to conclude that $J(R)^n = 0$. In fact, $J(R)$ *is* finitely generated; this follows from the Hopkins-Levitzky Theorem. But, since we didn't prove it, and its proof uses the Theorem we are trying to prove, we find a work-around.

Suppose that $J(R)^n = J(R)^{n+1}$, but $J(R)^n \neq 0$. Then

$$J(R)^n J(R) \neq 0.$$

We then consider a minimal non-zero left ideal M such that $J(R)^n M \neq 0$. Such an ideal exists by artinian induction. Thus, there is some $a \in M$ such that $J(R)^n a \neq 0$ and so $J(R)^n Ra \neq 0$. The minimality of M implies that in fact $M = Ra$ and, in particular, is finitely generated. Moreover

$$J(R)^n(J(R)M) = J(R)^{n+1}M = J(R)^n M \neq 0.$$

Note that $J(R)M$ is also a left ideal. As $J(R)^n(J(R)M) \neq 0$ and $M \supseteq J(R)M$, the minimality of M implies that $M = J(R)M$. As M is finitely generated we can apply Nakayama's Lemma and deduce that $M = 0$. Contradiction. □

27. Jacobson's density theorem

27.1. **Preparations.**

[39] If I and J are left ideals then $IJ = \{\sum a_i b_i : a_i \in I, b_i \in J\}$ is a left ideal contained in J. In particular, it makes sense to talk about I^n for any left ideal I and we have $I^n \supseteq I^{n+1}$.

27.1.1. *Faithful modules.* Let R be a ring and let $M \in {}_{\mathbf{R}}\mathbf{Mod}$ be a left R-module. Recall that M is called **faithful** if for $r \in R$,

$$rm = 0, \forall m \in M \quad \Rightarrow \quad r = 0.$$

In words, a non-zero element of R cannot act as zero on M. Recall also the **annihilator** $\mathrm{Ann}(M)$ of a module M. It is a two-sided ideal of R defined as

$$\mathrm{Ann}(M) := \{r \in R : rm = 0, \forall m \in M\}.$$

Thus,

$$M \text{ is faithful} \quad \Leftrightarrow \quad \mathrm{Ann}(M) = \{0\}.$$

27.1.2. *Dense subrings.* Let D be a division ring and let V be a left D-module. Let R be a ring contained in $\mathrm{End}_D(V)$. We say that R is **dense** in $\mathrm{End}_D(V)$ if for every finitely-generated submodule $U \subseteq V$ and every $T \in \mathrm{End}_D(V)$ there is an $r \in R$ such that

$$r|_U = T|_U.$$

Equivalently, given any $T \in \mathrm{End}_D(V)$ and elements $v_1, \ldots, v_n \in V$ there is an $r \in R$ such that

$$r(v_1) = T(v_1), \ldots, r(v_n) = T(v_n).$$

Note that if V is finite-dimensional over D then R is dense in $\mathrm{End}_D(V)$ if and only in $R = \mathrm{End}_D(V)$. Thus, the notion is useful either when V has infinite dimension over D or is not known a priori to have finite dimension.

27.1.3. *The setting.* Let M be a simple R-module. Then $D = \mathrm{End}_R(M)$ is a division ring. Indeed, by Schur's lemma every non-zero element of D is an isomorphism and so has a two-sided inverse, which is just the inverse function. We have then, and this is an important observation,

$$M \in {}_{\mathbf{D}}\mathbf{Mod},$$

where for $T \in D$, $m \in M$, the module action of T on m (denoted Tm) is simply $Tm = T(m)$, evaluating the homomorphism T on the element m. Now, any element $r \in R$ defines a map

$$[r] \colon M \to M, \quad [r](m) = rm.$$

This map is additive and for $T \in D$ we have $T([r]m) = T(rm) = rT(m) = [r]T(m)$. The conclusion is that there is a natural homomorphism of rings,

$$R \to \mathrm{End}_D(M).$$

That is, multiplication-by-a-scalar commutes with endomorphisms of M, clearly, but the point is that viewing multiplication-by-a-scalar as a map $M \to M$ it becomes an endomorphism of M as a D-module. Furthermore, if M is a faithful R-module, then multiplication by $r \neq 0$ is never the zero map, so:

Corollary 27.1.1. *Let M be an R-module then the following holds:*

$$M \text{ faithful } R\text{-module} \implies R \hookrightarrow \mathrm{End}_D(M), \quad \text{where } D = \mathrm{End}_R(M),$$

and when M is simple, D is a division ring.

27.2. **Jacobson's Density Theorem.**

Theorem 27.2.1. (Jacobson) *Let R be a ring and let $M \in {}_{\mathbf{R}}\mathbf{Mod}$ be a simple and faithful R-module. Then R in dense in $\mathrm{End}_D(M)$.*

Before proving the theorem let's discuss its assumptions. To say M is simple is to say that $M \cong R/I$ as an R-module, where I is a maximal left ideal. The annihilator of M is thus

$$\mathrm{Ann}(M) = \{r \in R : rs \in I, \forall s \in R\}.$$

Note that $\mathrm{Ann}(M) \subset I$ (take $s = 1$). If I contains a non-zero right ideal J then for a non-zero $r \in J$ we would have $rs \in J \subset I, \forall s \in R$. That is, it would follow that M is not faithful. Conversely, if M is not faithful, then $\mathrm{Ann}(M)$ is a non-zero two-sided (hence, right) ideal contained in I. Thus, the faithful simple modules come from[40] maximal left ideals whose only sub right ideals are $\{0\}$. This is a very non-commutative situation!

Proof. We are given

$$v_1, \dots, v_n \in M,$$

that we may assume to be linearly independent over D (any finitely generated D-module of M is spanned by such) and any

$$u_1, \dots, u_n \in M.$$

We need to show

$$\exists r \in R, \quad rv_i = u_i, \qquad \forall i = 1, \dots, n.$$

Note that is not even necessary for the proof to assume that there is some $T \in \mathrm{End}_D(M)$ such that $Tv_i = u_i$, but that is true by "vector spaces theory".

Reduction step: It is enough to show

$$\exists r \in R, \quad rv_1 = \cdots = rv_{n-1} = 0,$$
$$rv_n \neq 0.$$

Indeed, if $rv_n \neq 0$ then Rrv_n is a non-zero left R-submodule of M and so $Rrv_n = M$. Thus, we can change r so that $rv_i = 0, i = 1, \dots, n-1$ and $rv_n = u_n$. Denote this r by r_n. By the same argument, for every i we have an element $r_i \in R$, such that

$$r_i v_j = \delta_{ij} u_i.$$

Take $r = r_1 + \cdots + r_n$. It satisfies $rv_i = u_i, \forall i$.

Now, suppose to the contrary that we have the implication

$$\forall r \in R, \ rv_1 = \cdots = rv_{n-1} = 0 \ \Rightarrow \ rv_n = 0.$$

Our strategy would be to construct R-module endomorphisms $h_1, \dots, h_{n-1}$ of M such that

$$v_n = h_1(v_1) + \cdots + h_{n-1}(v_{n-1}).$$

That would imply that $v_1, \dots, v_n$ are linearly dependent over D and so lead to a contradiction.

Let $w = (v_1, \dots, v_{n-1}) \in M^{n-1}$. Then $Rw \subseteq M^{n-1}$ is a submodule. Consider the map

$$f \colon Rw \to M, \quad f(rw) = rv_n.$$

[40] We say "come from" and not "correspond to" because it is possible that for two distinct ideals I_1, I_2 we have $R/I_1 \cong R/I_2$ as left R-modules. For example, take $R = M_n(D)$ and the ideals I_1 =first colum is zero, I_2 = second column is zero. In both case $M_n(D)/I_j \cong D^n$ as left D-modules.

The map f is well defined: If $rw = r'w$ then $(r-r')w = 0$ and so $(r-r')v_n = 0$ by our assumption, meaning $rv_n = r'v_n$. It is easy to verify now that f is a homomorphisms of R-modules. As M is simple, M^{n-1} is semi-simple and thus there is an R-module P such that

$$M^{n-1} = Rw \oplus P.$$

We can therefore extend f to M^{n-1} by 0 on P. Using the same notation, we got an R-module homomorphism

$$f\colon M^{n-1} \to M,$$

such that $f(rw) = rv_n$. Using it, we define homomorphisms of R-modules

$$h_i : M \to M, \quad h_i(m) = f(0,\dots,\underset{i}{m},\dots,0).$$

Then,

$$\begin{aligned} v_n &= f(w) \\ &= f(v_1,\dots,v_{n-1}) \\ &= f(v_1,0,\dots,0) + f(0,v_2,0,\dots,0) + \cdots + f(0,\dots,0,v_{n-1}) \\ &= h_1(v_1) + h_2(v_2) + \cdots + h_{n-1}(v_{n-1}). \end{aligned}$$

□

28. Structure of artinian rings

28.1. **Simple rings.** A ring R is called **simple** if it has no 2-sided ideals except for $\{0\}$ and R. For example, any division ring is a simple ring because it doesn't even have non-trivial left ideals. One can prove that if D is a division ring then $M_n(D)$ is a simple ring. Note though that the ring $M_n(D)$ is not simple as a module over itself for $n > 1$; in fact it is the sum of n simple submodules, as we have already seen. Thus, a simple ring is not a simple module over itself and some care has to be taken regarding terminology. That said, if R is a simple ring then $J(R) = \{0\}$ and so if R is artinian, R is at least semi-simple (namely, as a module over itself). See Proposition 25.2.2. But, in fact, we can say much more.

Theorem 28.1.1. (Wedderburn) *Let R be a simple artinian ring. Then $R \cong M_n(D)$ for some division ring D.*

Proof. Let I be a maximal left ideal of R and let $M = R/I$. Then $\mathrm{Ann}(M) \subset I$ is a proper two-sided ideal, hence $\{0\}$. It follows that M is a simple faithful R-module. By Jacobson's Density Theorem, $R \subseteq \mathrm{End}_D(M)$ is dense, where $D = \mathrm{End}_R(M)$ is a division ring. If is enough to show that M is a finite-dimensional D-module. Indeed, in this case:

- $R = \mathrm{End}_D(M)$.
- $\mathrm{End}_D(M) = M_n(D)$, where $n = \dim_D(M)$.[41]

Suppose that M is infinite-dimensional over D. Let $a_1, a_2, a_3, \dots$ be elements of M that are linearly independent over D. With the notation

$$\mathrm{Ann}_R(a_1,\dots,a_n) = \{r \in R : ra_i = 0, i = 1,\dots,n\},$$

we obtain a sequence of decreasing left ideals of R,

$$\mathrm{Ann}_R(a_1) \supseteq \mathrm{Ann}(a_1,a_2) \supseteq \mathrm{Ann}_R(a_1,a_2,a_3) \supseteq \dots$$

[41] This is when we let $M_n(D)$ act from the right. If we want $M_n(D)$ to act from the left, as R does, then the statement is $R = M_n(D)^{\mathrm{op}}$. Here $M_n(D)^{\mathrm{op}}$ is the **opposite ring**; it is the same set as $M_n(D)$ with the same addition, but where multiplication is now defined as $A * B := BA$. If we define now for $v \in D^n$, thought of as a row vector, $A * v := vA$ then we have $(A*B)*v = v(A*B) = vBA = A*(vB) = A*(B*v)$ and so we get a right action of $M_n(D)$ on D^n. Note though that $M_n(D)^{\mathrm{op}} \cong M_n(D)$ as rings by the map $M \mapsto {}^tM$.

Since R is artinian, for some $n \geq 1$ we must have

$$\mathrm{Ann}_R(a_1,\dots,a_n) = \mathrm{Ann}_R(a_1,\dots,a_n,a_{n+1}),$$

which translates to the statement: if $r \in R$ is such that $ra_1 = \cdots = ra_n = 0$ then also $ra_{n+1} = 0$. However, using the independence over D, we can find an endomorphism T of M as a module over D such that

$$Ta_1 = \cdots = Ta_n = 0, \quad Ta_{n+1} \neq 0.$$

(Complete the $\{a_i\}$ to a basis over D; one can specify a map anyway one wants on a basis.) Since R is dense in $\mathrm{End}_D(M)$, there is some element $r \in R$ such that $Ta_i = ra_i, i = 1,\dots,n+1$ and in particular $ra_1 = \cdots = ra_n = 0$ and $ra_{n+1} \neq 0$. Contradiction. □

Note that during the proof we proved the following.

Lemma 28.1.2. *Let R' be an artinian ring, D' a division algebra and M' a D'-module. Suppose that $R' \subseteq \mathrm{End}_{D'}(M')$ and dense. Then $\dim_{D'}(M') < \infty$ and hence $R' = \mathrm{End}_{D'}(M')$.*

28.2. **The Artin-Wedderburn Theorem.**

Theorem 28.2.1. (Artin-Wedderburn) *Let R be a non-zero semisimple artinian ring. Then,*

$$R \cong M_{n_1}(D_1) \times \cdots \times M_{n_k}(D_k),$$

where each D_i is a division ring and the n_i are positive integers.

Proof. Since R is artinian and semisimple, we know that

$$R = M_1 \oplus \cdots \oplus M_\ell,$$

a finite direct sum of simple R-modules (Propositions 25.2.1 and 25.2.2). Here each M_i is a minimal left ideal of R. Note that this is not a decomposition of rings, merely of R-modules, so one cannot say that multiplication happens component-wise. Still, for $1 \leq i \leq \ell$, let

$$I_i = \oplus_{j \neq i} M_j.$$

Then I_i is a left R-ideal and $R/I_i \cong M_i$ and so I_i is a maximal ideal of R. We see this way too that $J(R) = \{0\}$ as already $\cap_{i=1}^{\ell} I_i = \{0\}$. At any rate, $M_i = R/I_i$ is a simple R-module that is simple and faithful over the artinian ring $R/\operatorname{Ann}(M_i)$. We apply Lemma 28.1.2 to $R' = R/\operatorname{Ann}(M_i)$, $M' = M_i$, $D' = \mathrm{End}_{R'}(M') =: D_i$, which is a division ring. The conditions of the Lemma hold by Jacobson's Density Theorem and we thus conclude that

$$R/\operatorname{Ann}(M_i) \cong \mathrm{End}_{D_i}(R/I_i) \cong M_{n_i}(D_i),$$

where $D_i = \mathrm{End}_{R/\operatorname{Ann}(M_i)}(M_i)$ is a division ring and $n_i = \dim_{D_i}(M_i)$.

We have a ring homomorphism

$$R \to R/\operatorname{Ann}(M_1) \times \cdots \times R/\operatorname{Ann}(M_\ell),$$

which is injective because $\cap_{i=1}^{\ell} \operatorname{Ann}(M_i) = \cap_{i=1}^{\ell} \operatorname{Ann}(R/I_i) \subseteq \cap_{i=1}^{\ell} \operatorname{Ann}(I_i) = \{0\}$. Thus, the following lemma completes the proof:

Lemma 28.2.2. *Let R be a ring and $R \hookrightarrow R_1 \times \cdots \times R_\ell$ an injective ring homomorphism. Suppose that each R_i is a simple ring and that for every i the composition with the projection π_i on the i-th component*

$$R \longrightarrow R_1 \times \cdots \times R_\ell \xrightarrow{\pi_i} R_i$$

is surjective. Then R is isomorphic to a sub-product of $R_1 \times \cdots \times R_\ell$. More precisely, there is a subset $I \subseteq \{1,2,\ldots,\ell\}$ such that the composition

$$R \longrightarrow R_1 \times \cdots \times R_\ell \xrightarrow{\pi_I} R_I$$

is an isomorphism, where $R_I = \prod_{i\in I} R_i$ and $\pi_I(r_i)_{i=1}^{\ell} = (r_i)_{i\in i}$ is projection on the coordinates in I.

Proof. (Lemma) We prove the lemma by induction on ℓ, where the case $\ell = 1$ is clear.

For $\ell > 1$, let $S = R_2 \times \cdots \times R_\ell$. Let $\pi_S \colon R \to S$ the natural map. $\pi_S(R) \subseteq S$ and so, by induction, $\pi_S(R)$ is isomorphic to a sub-product of S via the natural projection; that is, there is a subset J of $\{2,\ldots,\ell\}$ such that via the natural projection π_J the ring $\pi_S(R)$ is isomorphic to $\prod_{j\in J} R_j$. Note that this forces that $\pi_S(R) \hookrightarrow \prod_{j\in J} R_j$; we may therefore assume that π_S is surjective to begin with. So, we arrive at the situation:

$$R \hookrightarrow R_1 \times S, \quad \pi_S \colon R \twoheadrightarrow S, \quad \pi_1 \colon R \twoheadrightarrow R_1,$$

and we want to prove that either $R = R_1 \times S$, or that R is isomorphic to S via π_S.[42]

Let

$$I = \{x \in R_1 : (x,0) \in R\} = \pi_1(\mathbf{Ker}(\pi_S)).$$

Note that I is a two-sided ideal of R_1: if $x \in I$ and $t \in R_1$ there is some $y \in S$ such that $(t,y) \in R$ (π_1 is surjective). Then $(xt,0) = (x,0)(t,y) \in R$, implying $xt \in I$, and similarly for tx.

Since R_1 is simple there are two options:

(1) Either $I = 0$, in which case the map π_S is also injective, $R \cong S$ via π_S, and we are done. Or,
(2) $I = R_1$, in which case $\mathbf{Ker}(\pi_S) = R_1 \times \{0\}$ and both R and $R_1 \times S$ surject onto S with the same kernel and so they are equal: $R = R_1 \times S$.

□

□

29. Central simple algebras and the Brauer group

Our discussion in this section follows Jacobson's book *Basic Algebra II*.

Let F be a field and let A be an F-algebra which means that F is contained in the centre of A – it is a subring of A and each element of it commutes with any element of A. Assume further that A is finite dimensional over F, and hence an artinian ring. If A is a simple algebra, then by Theorem 28.1.1, we have

$$A \cong M_n(D),$$

for some division ring D and an integer $n \geq 1$. It is not hard to check that A and D have the same centre, say K, which is thus a field containing F. We can consider A as a K-algebra too and then K is its centre.

In general, a K-algebra A with the property that K is precisely its centre is called a **central K-algebra**. Thus, the discussion above explains that a simple finite dimensional F-algebra is always a central simple K-algebra for some field K containing F. Further, K is the centre of D under the isomorphism $A \cong M_n(D)$. The simplest examples of central simple K-algebras of finite dimension are thus $M_n(K)$ for $n = 1,2,\ldots$.

In this chapter we will consider the collection of central simple finite dimensional K-algebras, up to a certain equivalence relation, and some of the remarkable properties of this set.

[42] Not necessarily equal to S! A good example to keep in mind is $\mathbb{F}_p \subset \mathbb{F}_p \times \mathbb{F}_p$, diagonally. Also, remark that in the second case when R is isomorphic to S, *if S is simple*, reversing the role of R_1 and S we will also find that R is isomorphic to R_1, by π_1.

29.1. **Tensor product of K-algebras.** Let A and B be central K-algebras of finite dimension over a field K. Then

$$A \otimes_K B$$

is a K-algebra (§ 2.4) and if $\{v_i\}$ is a K-basis for A and $\{w_j\}$ a K-basis for B then $\{v_i \otimes w_j\}$ is a K-basis for $A \otimes_K B$ and, in particular,

$$\dim_K(A \otimes_K B) = \dim_K(A) \cdot \dim_K(B).$$

Note that we can view A and B as subalgebras of $A \otimes_K B$ by

$$A \hookrightarrow A \otimes_K B, \quad a \mapsto a \otimes 1; \qquad B \hookrightarrow A \otimes_K B, \quad b \mapsto 1 \otimes b.$$

Proposition 29.1.1. *Let A, B be sub K-algebras of a K-algebra C, all finite dimensional over K. Then*

$$C \cong A \otimes_K B,$$

as K-algebras, if and only if the following conditions hold:

(1) $ab = ba, \forall a \in A, b \in B$.
(2) $C = AB$ (meaning, any element of C has the form $\sum a_i b_i$ for some $a_i \in A, b_i \in B$).
(3) $\dim_K(C) = \dim_K(A) \cdot \dim_K(B)$.

Proof. The only if part is straightforward. For the if part, note that by the usual argument there is a homomorphism of K-modules

$$A \otimes_K B \to C, \quad a \otimes b \mapsto ab.$$

The first condition guarantees that this is a ring homomorphism as well, the second condition that the map is surjective and, from the theory of vector spaces, the third condition guarantees it is injective. □

Corollary 29.1.2. *If R is a finite dimensional K-algebra,*[43]

$$M_n(K) \otimes_K R \cong M_n(R).$$

Proof. This follows from the proposition by letting $A = M_n(K)$ and $B = R \cdot I_n = \{\mathrm{diag}(r, \dots, r) : r \in R\}$. □

Corollary 29.1.3. *Let K be a field then*

$$M_n(K) \otimes_K M_m(K) \cong M_{mn}(K).$$

Proof. From the previous corollary, $M_n(K) \otimes_K M_m(K) \cong M_n(M_m(K))$ ($n \times n$ matrices whose entries are $m \times m$ matrices). A direct verification shows that $M_n(M_m(K)) \cong M_{mn}(K)$, essentially by removing the brackets around the matrices appearing as entries in $M_n(M_m(K))$. □

If A is a K-algebra, we denote by A^{op} the **opposite algebra**. It is equal to A as an abelian group, but we define a product $a * b$ by $a * b = ba$, where ba is the original product, in A, of b and a. It is again a K-algebra, because as K is commutative $K^{\mathrm{op}} = K$. We define

$$A^e = A \otimes_K A^{\mathrm{op}},$$

and, following Jacobson, call it the **enveloping algebra** of A. Note that multiplication is given on pure tensors by

$$a_1 \otimes \alpha_1 \cdot a_2 \otimes \alpha_2 = a_1 a_2 \otimes \alpha_2 \alpha_1.$$

If A is a sub K-algebra of a K-algebra B, then B has a left action of A^e given on pure tensors by

$$a \otimes \alpha \cdot b = ab\alpha.$$

[43] This is true also without the finite-dimensional hypothesis.

Indeed, $a_1 \otimes \alpha_1 \cdot (a_2 \otimes \alpha_2 \cdot b) = a_1 \otimes \alpha_1 \cdot a_2 b \alpha_2 = a_1 a_2 b \alpha_2 \alpha_1 = a_1 a_2 \otimes \alpha_2 \alpha_1 \cdot b = (a_1 \otimes \alpha_1 \cdot a_2 \otimes \alpha_2) b$, which explains why we define the multiplication in A^e the way we do.

Taking the particular case $A = B$ we find that A is a left A^e-module. It's another way to say that A is an A-A bimodule. A submodule of A (as an A^e-module) is a two-sided ideal of A. Thus, if A is a simple K-algebra, its only submodules over A^e are $\{0\}$ and A, namely, it becomes a simple A^e module. This proves the first part of the following lemma.

Lemma 29.1.4. *Let A be a central simple K-algebra.*

(1) A is a simple A^e-module.
(2) $\mathrm{End}_{A^e}(A) = K$.

Proof. Let R be any ring. Viewing R as a left R-module, we have $\mathrm{End}_R(R) = R^{\mathrm{op}}$, by sending $f \in \mathrm{End}_R(R)$ to $f(1)$ and, conversely, given $r \in R$ we define $f(x) = xr$. Viewing R as a right R-module (equivalently(!), as a left R^{op}-module) we have the same identification $\mathrm{End}_R(R) = R$, where to $r \in R$ we now associate the map $f(x) = rx$.

Now, a homomorphism $f\colon A \to A$ as an A^e-module, is both a homomorphism of A as an $A = A \otimes 1$-module and thus of the form $x \mapsto xr_1$ and a homomorphism of A as an $A^{\mathrm{op}} = 1 \otimes A^{\mathrm{op}}$-module and thus of the form $x \mapsto r_2 x$. Taking $x = 1$ we find that $r_1 = r_2$ and then that for all $x \in R$, $xr_1 = r_1 x$, which means that r_1 is in the centre of A. That is, $r_1 \in K$. □

Theorem 29.1.5. *Let A be a finite-dimensional central simple K-algebra, $\dim_K(A) = n$. Then*

$$A^e = A \otimes_K A^{\mathrm{op}} \cong M_n(K).$$

Proof. As noted, the algebra A is a simple A^e-module. We don't know a priori that it is a faithful A^e module, but it is simple and faithful over $B := A^e/I$, where I is the two-sided ideal of A^e that is the annihilator of A as an A^e-module. By Jacobson's density theorem, using that $\mathrm{End}_{A^e}(A) = \mathrm{End}_B(A) = K$, we have that $B \hookrightarrow \mathrm{End}_K(A) \cong \mathrm{End}_K(K^n) \cong M_n(K)$ as a dense subring. And, since B contains K and dense, we actually have $B \cong \mathrm{End}_K(A) \cong M_n(K)$. Thus, we get a surjective ring homomorphism $A^e \to M_n(K)$, and comparing the dimensions of both sides as K-vector spaces, we find that this must be an isomorphism. □

Example 29.1.6. Suppose that K is a field of characteristic different than 2. Then any quaternion algebra B over K is a central simple K-algebra and has a presentation

$$K \oplus Ki \oplus Kj \oplus Kk,$$

for some constants $\alpha, \beta \in K$, where

$$i^2 = \alpha, \quad j^2 = \beta, \quad ij = k, \quad ij = -ji.$$

We previously denoted this quaternion algebra $\left(\frac{\alpha,\beta}{k}\right)$. Using this, one can verify that the canonical involution of B, namely, the map

$$a = x + yi + zj + wk \;\; \mapsto \;\; \bar{a} = \mathrm{Tr}(a) - a = x - yi - zj - wk,$$

provides an isomorphism $B \cong B^{\mathrm{op}}$. The key identities are $\overline{a+b} = \bar{a} + \bar{b}, \overline{ab} = \bar{b}\bar{a}$. We therefore conclude that

$$B \otimes_K B \cong M_4(K).$$

This is not an obvious statement at all! For example, the right hand side has a collection of 16 special elements – the elementary matrices E_{ij} that satisfy the relations $E_{ij}E_{k\ell} = \delta_{jk}E_{i\ell}$. How can we write them as elements of the left hand side?

29.2. **Tensor product of central simple K-algebras.** Our goal in this section is to prove the following statement: *Let $A_1,\dots,A_n$ be finite-dimensional central simple K-algebra then also the tensor product $A_1\otimes_K A_2\otimes\cdots\otimes A_n$ is a finite dimensional central simple K-algebra.* But this is not an easy statement to prove and we will have to build up towards it.

As we will deal a lot with finite-dimensional central simple K-algebras A, we will simply say that A is an **fdCS** K-algebra.

For a K-subalgebra A of a K-algebra B, let

$$C_B(A)=\{b\in B : ba=ab, \forall a\in A\};$$

it is a K-subalgebra of B called the **centralizer** of A in B.

Lemma 29.2.1. *Let A be an fdCS K-subalgebra of a finite dimensional K-algebra B. Let $C=C_B(A)$ be the centralizer of A in B. Then,*

$$B\cong A\otimes_K C,$$

and the map

$$I\mapsto AI,$$

is a bijection between the two-sided ideals of C and the two-sided ideals of B. Moreover, the centre of B coincides with the centre of C.

Proof. By Theorem 29.1.5, $A^e\cong M_n(K)$, where $n=\dim_K(A)$ and thus A^e is a simple artinian ring with a unique simple module up to isomorphism. By Lemma 29.1.4, A itself is a simple A^e-module. Note that A, as an A^e-module, has a special generator c, namely $c=1$, with the following properties:

(1) c generates A as an $A=A\otimes 1\subset A^e$-module,
(2) $(a\otimes 1)\cdot c=(1\otimes a)\cdot c$, and
(3) if $(a\otimes 1)c=0$ then $a=0$.

Thus, *any* simple A^e-module has a generator c with the same properties.

As A^e is also artinian ring, being of finite dimension over K, it is semi-simple (Theorem 28.1.1, or simply from the isomorphism with $M_n(K)$) and any module over it is isomorphic to a sum of copies of A.[44] In particular, B itself, considered as A^e-module, is a sum of copies of A. Thus,

$$B=\oplus_{\alpha\in I}Ac_\alpha,$$

where c_α are elements of B that are generators with the said properties, and one of them can be chosen to be 1; say $c_{\alpha_0}=1$. It therefore follows that $c_\alpha\in C$ and any element of B can be written uniquely as $\sum_\alpha a_\alpha c_\alpha$, with $a_\alpha\in A$. A computation based on checking when does an element of the form $\sum_\alpha a_\alpha c_\alpha$ commutes with ac_{α_0} for any $a\in A$, yields that

$$C=\{\sum_\alpha k_\alpha c_\alpha : k_\alpha\in K\}.$$

It follows from Proposition 29.1.1 that

$$B\cong A\otimes_K C.$$

Note that the centre of B commutes with A and hence is contained in C. But then it commutes with the elements of C too, so it is contained in the centre of C. But $B=AC$, and C commutes with A and so the centre of C commutes with every element of B and we find that B and C have the same centre.

It remains to prove the statement concerning ideals. Let I be a two-sided ideal of C then $AI=\{\sum a_\alpha i_\alpha : a_\alpha\in A, i_\alpha\in I\}$ is a two-sided ideal of B, using that $B=AC$ and C commutes with A. In the tensor product

[44] In fact, since we know that $A^e\cong M_n(K)$, we can deduce all its properties rather directly instead of appealing to the general theory.

presentation $B = A \otimes_K C$, AI is the ideal

$$\left\{ \sum_\alpha a_\alpha \otimes i_\alpha : a_\alpha \in A, i_\alpha \in I \right\}. \tag{31}$$

Let us choose a basis for A over K, say $\{x_1 = 1, \ldots, x_n\}$. Then any element in $A \otimes_K C$ has a *unique* expression as

$$\sum_{i=1}^n x_i \otimes d_i, \quad d_i \in C,$$

and so every element of B has a *unique* expression as $\sum_{i=1}^n x_i d_i$; C itself are the elements $x_1 d_1, d_1 \in C$. By expanding the a_α appearing in (31) according to the basis $\{x_i\}$, we find that the ideal AI is precisely

$$\{\sum_{i=1}^n x_i \otimes d_i, \quad d_i \in I\}.$$

We see then that $AI \cap C = I$. Thus, we have an injective map from two-sided ideals of C to two-sided ideals of B. We need to show that every two-sided ideal of B has this form.

Let J be a two-sided ideal of B. Let $I = J \cap C$, which is a two-sided ideal of C. As J is an A^e-submodule of B, $J = \sum_\alpha A d_\alpha$ for some special generators $d_\alpha \in B$. But property (2) of a special generator implies that $d_\alpha \in C$, for all α. So, in fact, $d_\alpha \in J \cap C = I$ and it follows that $J = AI$. □

Corollary 29.2.2. *Let A be an fdCS K-algebra and let C be any K-algebra of finite dimension over K. Then the map*

$$I \mapsto A \otimes_K I,$$

is a bijection between two-sided ideals of C and two-sided ideals of $A \otimes_K C$. Moreover, the centre of $A \otimes_K C$ is the centre of C, under the identification $C = 1 \otimes C$.

Proof. Let $\{y_1 = 1, \ldots, y_m\}$ be a basis for C over K. Then any element of $A \otimes_K C$ can be written uniquely as $\sum_{i=1}^m a_i \otimes y_i, a_i \in A$. An element of A, written as $a \otimes y_1$, commutes with such a sum if and only if $a_i \in K$ for all i (because $(a \otimes 1)(\sum_{i=1}^m a_1 \otimes y_i) = \sum_{i=1}^m a a_i \otimes y_i$, etc.). Thus, the centralizer of A is $\{\sum_{i=1}^m a_i \otimes y_i : a_i \in K\} = C$. We can therefore apply Lemma 29.2.1 to $B = A \otimes_K C$. □

It follows directly from Lemma 29.2.1 that the following holds true.

Corollary 29.2.3. *Let A be an fdCS K-algebra and C a finite-dimensional K-algebra. Then*

$$A \otimes_K C \text{ is central over } K \iff C \text{ is central over } K,$$

and

$$A \otimes_K C \text{ is a simple } K \text{ algebra} \iff C \text{ is a simple } K \text{ algebra.}$$

By iterating, we find the following corollary that will be very important for the theory of the Brauer group.

Corollary 29.2.4. *Let $A_1, \ldots, A_n$ be fdCS K-algebras then also $A_1 \otimes_K A_2 \otimes_K \cdots \otimes_K A_n$ is a fdCS K-algebra.*

29.3. **The double-centralizer theorem.** Recall the Jacobson Density Theorem: *Let R be a ring and let $M \in {}_{\mathbf{R}}\mathbf{Mod}$ be a simple and faithful R-module. Then R in dense in $\mathrm{End}_D(M)$, where $D = \mathrm{End}_R(M)$.*

We make some remarks about this theorem. First, the assumption that M is faithful is not that important. Dropping this assumption, one may derive by the same proof that the image of R in $\mathrm{End}_D(M)$ is dense in it. Secondly, it turns out that the theorem also holds when M is assumed to be semi-simple. We do not give the proof here, but it can be found, for example, in Jacobson's Basic Algebra II, §4.3. The statement is this:

Theorem 29.3.1. (Jacobson) *Let R be a ring and $M \in {}_{\mathbf{R}}\mathbf{Mod}$ a semisimple R-module. Then the image of R is dense in $\mathrm{End}_D(M)$, where $D = \mathrm{End}_R(M)$.*

One application of this stronger statement is the following very useful theorem. For the proof see loc. cit. §4.6.

Theorem 29.3.2. (Double Centralizer Theorem) *Let A be a semisimple K-sub algebra of an fdCS algebra B. Then the double centralizer satisfies*

$$C_B(C_B(A)) = A.$$

This theorem, plus substantial additional work, allows one to deduce the following statement (loc. cit.):

Theorem 29.3.3. *Let A be a simple K-subalgebra of an fdCS K-algebra B. Then,*

$$\dim_K(B) = \dim_K(A) \cdot \dim_K(C_B(A)).$$

A very useful conclusion is the following:

Corollary 29.3.4. *Let A be a commutative subfield of the matrix algebra $M_n(K)$ that contains K. Then $\dim_K(A) | n$ and if $\dim_K(A) = n$ then $A = C_{M_n(K)}(A)$.*

Proof. Let $d := \dim_K(A)$. Note that $A \subseteq C_{M_n(K)}(A)$ and thus $d^2 | \dim_K(A) \cdot \dim_K(C_B(A)) = n^2$. It follows that $d | n$.

If $d = n$ then we must have $d = \dim_K(C_{M_n(K)}(A))$; in that case the inclusion $A \subseteq C_{M_n(K)}(A)$ must be an equality. □

Example 29.3.5. To illustrate the Theorem, we note that it implies that any subfield of a quaternion algebra over K is either K or a quadratic extension L of K. In the latter case, any element of the quaternion algebra that commutes with L is actually in L.

The corollary also implies that no quadratic field extension L of K can be embedded in $M_3(K)$. And so-on.

29.4. **The Brauer group.** Let K be a field and consider fdCS K-algebras A, B. We say that A is (Brauer) **similar** to B, and we write $A \sim B$, if there are some positive integers m, n such that

$$M_m(A) \cong M_n(B),$$

as K-algebras. Note that as $M_m(A) \cong M_m(K) \otimes_K A$, we can rephrase and say

$$A \sim B \quad \Leftrightarrow \quad \exists m, n \in \mathbb{N}^+,\ M_m(K) \otimes_K A \cong M_n(K) \otimes_K B.$$

This relation is clearly reflexive and symmetric. It is also transitive. Suppose $B \sim C$ and $M_s(B) \cong M_t(C)$. Then

$$M_{sm}(K) \otimes A \cong M_s(K) \otimes M_m(K) \otimes A \cong M_s(K) \otimes M_n(K) \otimes B \cong M_n(K) \otimes M_s(K) \otimes B$$
$$\cong M_n(K) \otimes M_t(K) \otimes C \cong M_{nt}(K) \otimes C.$$

Our main goal is to put a group structure on the equivalence classes of fdCS K-algebras. Suppose that $A \sim A_1$ and $B \sim B_1$, say $M_m(K) \otimes_K A \cong M_{m_1}(K) \otimes_K A_1$ and $M_n(K) \otimes_K B \cong M_{n_1}(K) \otimes_K B_1$. Note that,

$$M_{mn}(K) \underset{K}{\otimes} A \underset{K}{\otimes} B \cong M_m(K) \underset{K}{\otimes} M_n(K) \underset{K}{\otimes} A \underset{K}{\otimes} B \cong M_m(K) \underset{K}{\otimes} A \underset{K}{\otimes} M_n(K) \underset{K}{\otimes} B.$$

(where we used the general observation that for any K-algebras C, D one has $C \otimes_K D \cong D \otimes_K C$, induced from $c \otimes d \mapsto d \otimes c$, and applied it to $M_s(K)$ and $M_n(K)$). The group structure is defined for any two fdCS K-algebras, A, B, by defining the product of A and B as

$$A \otimes_K B.$$

Since

$$M_m(K) \underset{K}{\otimes} A \underset{K}{\otimes} M_n(K) \underset{K}{\otimes} B \cong M_{m_1}(K) \underset{K}{\otimes} A_1 \underset{K}{\otimes} M_{n_1}(K) \underset{K}{\otimes} B_1 \cong M_{m_1 n_1}(K) \underset{K}{\otimes} A_1 \underset{K}{\otimes} B_1,$$

we get a well defined operation on Brauer similarity classes. Thus, we have defined a multiplication on equivalence classes of fdCS K-algebra, which is commutative and associative. The identity element is simply K (note that it is equivalent to $M_n(K)$ for all $n \geq 1$). We also have an inverse since we proved in Theorem 29.1.5 that

$$A \underset{K}{\otimes} A^{\mathrm{op}} \cong M_n(K),$$

where $n = \dim_K(A)$. By Theorem 28.1.1 any fdCS K-algebra is similar to a division ring D whose centre is K. Altogether, we proved the following.

Theorem 29.4.1. *The set of similarity classes of fdCS K-algebras forms an abelian group under a group law induced from the tensor product of K-algebras. The identity element is the equivalence class of K and the inverse of A is A^{op}. This group is called the* **Brauer group** *of K. We denote it $\mathrm{Br}(K)$. Every similarity class contains a division algebra over K.*

Remark 29.4.2. In fact, one can prove that if $D_i, i = 1, 2$ are central division algebras over K then $D_1 \sim D_2$ if and only if D_1 is isomorphic to D_2 as K-algebras.

Example 29.4.3. If A is a quaternion algebra over K then $A \cong A^{\mathrm{op}}$ and so $A \otimes_K A \cong M_4(K)$ (see Example 29.1.6). That is, a quaternion algebra A over a field K is an element of order 2. It is a deep theorem of A. S. Merkurjev that if the characeristic of K is not 2 then the subgroup of elements of order 2 of $\mathrm{Br}(K)$ is *generated* by quaternion algebras. Sometimes, for instance when K is a number field, every element of order 2 is represented by a quaternion algebra. On the other hand, A. Kresch proved that the elements of order 2 of $\mathrm{Br}(K)$, where K is the function field of a complex algebraic threefold are not all represented by quaternion algebras, in general.

Determining the Brauer group of a field is a difficult problem. Here are a few fundamental results:

(1) If K is algebraically closed, $\mathrm{Br}(K)$ is the trivial group. This is clear once one has proved that every finite dimensional division K-algebra is equal to K.
(2) Every element in $\mathrm{Br}(K)$ is of finite order. This is not an easy statement, but once one proves that every division algebra is a crossed product and relates that to cohomology, the result follows immediately from homological algebra.

(3) By Theorem 28.1.1, every fdCS algebra A over K is isomorphic to $M_n(D)$ for some division K-algebra D. In fact, as already mentioned, one can show that D is uniquely determined up to isomorphism. Thus, we conclude that every equivalence class of fdCS K-algebras, i.e. every element of $\mathrm{Br}(K)$, is represented by a division K-algebra D, unique up to isomorphism. The main point is that the tensor product of central division algebras need not be a division algebra, but is at least Brauer-similar to a central division algebra.
(4) We have $\mathrm{Br}(\mathbb{R}) \cong \mathbb{Z}/2\mathbb{Z}$, where the non-trivial element is the Hamilton quaternions.
(5) One of the major achievement of algebraic number theory in the last century was the determination of $\mathrm{Br}(K)$ for K a number field. To fix idea, consider the case of $\mathbb{Q}$. For every prime p, the association $A \mapsto A \otimes_{\mathbb{Q}} \mathbb{Q}_p$ provides a group homomorphism

$$\mathrm{Br}(\mathbb{Q}) \to \mathrm{Br}(\mathbb{Q}_p).$$

We will use this notation also for $p = \infty$ and by $\mathbb{Q}_\infty$ we will mean $\mathbb{R}$. It is a theorem, that for every prime p

$$\mathrm{Br}(\mathbb{Q}_p) \cong \mathbb{Q}/\mathbb{Z},$$

and a deep theorem that there is an exact sequecne

$$0 \longrightarrow \mathrm{Br}(\mathbb{Q}) \longrightarrow \oplus_{p\leq\infty}\mathrm{Br}(\mathbb{Q}_p) \longrightarrow \mathbb{Q}/\mathbb{Z} \longrightarrow 0.$$

The second arrow is $A \mapsto (A \otimes_{\mathbb{Q}} \mathbb{Q}_p)_{p\leq\infty}$ and, identifying the Brauer group of $\mathbb{Q}_p$ with $\mathbb{Q}/\mathbb{Z}$, and of $\mathbb{R}$ with $\frac{1}{2}\mathbb{Z}/\mathbb{Z}$, the next arrow is $(a_p)_{p\leq\infty} \mapsto \sum_{p\leq\infty} a_p$.

This exact sequence contains a lot of information. For example, it implies that if B is a quaternion algebra over $\mathbb{Q}$ then for almost all p, $B \otimes_{\mathbb{Q}} \mathbb{Q}_p \cong M_2(\mathbb{Q}_p)$ and the number of primes (counting ∞ as a prime) such that $B \otimes_{\mathbb{Q}} \mathbb{Q}_p \ncong M_2(\mathbb{Q}_p)$ is even and determines the quaternion algebra up to isomorphism.

Part 5. INTRODUCTION TO HOMOLOGICAL ALGEBRA

30. Exactness properties of functors

30.1. **Abelian categories.** R-modules are an example of an **abelian category**. We can think about an abelian category as a category where for any two objects A, B and morphisms $f, g, \in \mathrm{Mor}(A, B)$ one has their sum $f + g \in \mathrm{Mor}(A, B)$, and this addition makes $\mathrm{Mor}(A, B)$ into an abelian group. Moreover, the addition is natural with respect to composition; kernels and co-kernels of morphisms exist; finite direct sums and finite direct products exist and are equal. Further, there is a zero object denoted 0. A functor between abelian categories is called **additive** if $F0 = 0$, $F(f + g) = Ff + Fg$ and $F(X \oplus Y) \cong F(X) \oplus F(Y)$.

Our definition is hardly a proper definition, as we have not defined what are kernels and co-kernels in a category. The following paragraph explains why we allow ourselves this cavalier approach.

The main example of an abelian category is the category of (left, or right) R-modules, where R is any ring (always with 1). It is almost the most general example. Mitchell's Embedding Theorem states that any small[45] abelian category $\mathbf{C}$ is equivalent to a full subcategory of the category of modules over some (not necessarily commutative) ring R.

For the following, we can fix ideas and think about the abelian category in question as either the category of R-modules over a ring R, or the category of complexes of modules over a ring R (to be described later). *The functors we will consider will always be additive functors.*

30.2. **Exactness properties of functors.** An additive covariant functor F between abelian categories is called:

- **left-exact** if
$$0 \to A \to B \to C \text{ exact} \quad \Rightarrow \quad 0 \to FA \to FB \to FC \text{ exact.}$$
- **right-exact** if
$$A \to B \to C \to 0 \text{ exact} \quad \Rightarrow \quad FA \to FB \to FC \to 0 \text{ exact.}$$
- **exact** if it is both left and right-exact. This is equivalent to
$$0 \to A \to B \to C \to 0 \text{ exact} \quad \Rightarrow \quad 0 \to FA \to FB \to FC \to 0 \text{ exact.}$$

An additive contracovariant functor F between abelian categories is called:

- **left-exact** if
$$A \to B \to C \to 0 \text{ exact} \quad \Rightarrow \quad 0 \to FC \to FB \to FA \text{ exact.}$$
- **right-exact** if
$$0 \to A \to B \to C \text{ exact} \quad \Rightarrow \quad FC \to FB \to FC \to 0 \text{ exact.}$$
- **exact** if it is both left and right-exact. This is equivalent to
$$0 \to A \to B \to C \to 0 \text{ exact} \quad \Rightarrow \quad 0 \to FC \to FB \to FA \to 0 \text{ exact.}$$

Thus, the use of "left" or "right" is according to the output of the functor.

Let R be a ring and $M \in \mathbf{Mod_R}$, $N \in {}_\mathbf{R}\mathbf{Mod}$. We may then consider the additive functors
$$M \otimes_R (\cdot) : {}_\mathbf{R}\mathbf{Mod} \to \mathbf{AbGps}, \quad (\cdot) \otimes_R N : \mathbf{Mod_R} \to \mathbf{AbGps}.$$

[45] That means that the collection of objects of C forms a set and for every two objects X, Y of C, $\mathrm{Hom}_C(X, Y)$ is a set.

Theorem 30.2.1. *$M \otimes_R (\cdot)$ and $(\cdot) \otimes_R N$ are right-exact functors.*

Proof. We prove the claim for the functor $M \otimes_R (\cdot)$; the other case is entirely similar. Let

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0.$$

be an exact sequence in ${}_R\mathbf{Mod}$. Then, from properties of tensor products and their functoriality, we have a **complex**, i.e. the image of any morphism is contained in the kernel of the next,

$$M \otimes A \xrightarrow{1\otimes\alpha} M \otimes B \xrightarrow{1\otimes\beta} M \otimes C \longrightarrow 0.$$

We need to show it is exact and that amounts to the following:

- $1 \otimes \beta$ *is surjective.* This is clear: given $\sum m_i \otimes c_i \in M \otimes C$ choose $b_i \in B$ such that $\beta(b_i) = c_i$. Then, $(1 \otimes \beta)(\sum m_i \otimes b_i) = \sum m_i \otimes c_i$.
- $(1 \otimes \beta) \circ (1 \otimes \alpha) = 0$. As said, this is clear from functoriality because $(1 \otimes \beta) \circ (1 \otimes \alpha) = 1 \otimes (\beta \circ \alpha) = 1 \otimes 0 = 0$.
- $\mathrm{Ker}(1 \otimes \beta) \subseteq \mathrm{Im}(1 \otimes \alpha)$. Let $E = \mathrm{Im}(1 \otimes \alpha)$. We already showed that $E \subseteq \mathrm{Ker}(1 \otimes \beta)$. By the first isomorphism theorem there is an induced surjective map

  $$\bar{\beta} : M \otimes B/E \to M \otimes C,$$

  with kernel $\mathrm{Ker}(1 \otimes \beta)/\mathrm{Im}(1 \otimes \alpha)$, and it is enough to prove $\bar{\beta}$ is an isomorphism (and we know already that it is surjective). We prove that by constructing the inverse map.

  Let $f : M \times C \to M \otimes B/E$ be the map

  $$f(m, c) = m \otimes b \pmod{E}, \quad \text{any } b \text{ such that } \beta(b) = c.$$

  We claim that f is well-defined: First, such a b exists because β is surjective. If also $\beta(b') = c$ then $\beta(b - b') = 0$ and thus there is some $a \in A$ such that $\alpha(a) = b - b'$. Then $m \otimes b - m \otimes b' = m \otimes (b - b') = (1 \otimes \alpha)(m \otimes a) \in E$.

  It is easy to check that f is R-bilinear and so we get a well-defined homomorphism

  $$\bar{f} : M \otimes C \to M \otimes B/E, \quad \bar{f}(\sum m_i \otimes c_i) = \sum m_i \otimes b_i \quad (f(b_i) = c_i).$$

  Now, $\bar{f} \circ \bar{\beta} = id$. To prove that it is enough to prove that $\bar{f}(\bar{\beta}(\sum m_i \otimes b_i)) = \sum m_i \otimes b_i$, which is clear because we can choose b_i as lifts of the $\beta(b_i)$. This prove that $\bar{\beta}$ is injective, hence an isomorphism.

□

Theorem 30.2.2. *Let M be an R-module (left, or right) then:*

- $\mathrm{Hom}_R(M, \cdot)$ *is a left-exact covariant functor;*
- $\mathrm{Hom}_R(\cdot, M)$ *is left-exact contravariant functor.*

Proof. We assume that we are dealing with left R-modules, as the other case is the same. We prove the first statement, leaving the second as an exercise.

We need to show that if

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

is an exact sequence of modules then

$$0 \to \mathrm{Hom}_R(M, A) \to \mathrm{Hom}_R(M, B) \to \mathrm{Hom}_R(M, C)$$

is an exact sequence of abelian groups. We already know it is a complex. Thus, we need to show the following:

(1) That $\mathrm{Hom}_R(M,A) \to \mathrm{Hom}_R(M,B)$ is injective. Suppose that that $h \in \mathrm{Hom}_R(M,A)$ and $f \circ h = 0$. That is, for all $m \in M$, $f(h(m)) = 0$. Since f is injective, this implies that $h(m) = 0$ for all m and so that $h = 0$.
(2) Suppose that $h \in \mathrm{Hom}_R(M,B)$ and $g \circ h = 0$. Then, for all $m \in M$ we have $g(h(m)) = 0$ and hence $h(m) \in \mathrm{Ker}(g) = \mathrm{Im}(f)$. Thus, for all $m \in M$, $h(m) = f(a_m)$ for a unique $a_m \in A$ and one verifies, using uniqueness, that
$$M \to A, \quad m \mapsto a_m$$
is a homomorphism of R-modules; denote it h_1. Then, by definition, $h = f \circ h_1$.

□

This result suggests singling out those modules M with an additional exactness property. Let M be an R-module.

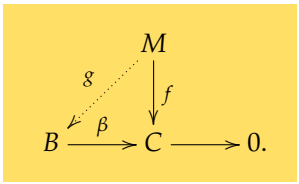
- M is called **projective** if $\mathrm{Hom}_R(M,\cdot)$ is exact;
- M is called **injective** if $\mathrm{Hom}_R(\cdot,M)$ is exact.

31. Projective modules

31.1. **Definition and basic properties.** Let R be a ring and M a left R-module. The notion of projective has to do with *lifting* homomorphisms *from* M.

Theorem 31.1.1. *M is projective if and only if given any diagram of R-modules and R-module homomorphisms, with β surjective, there exists a homomorphism g making it commutative:*

(32)



Proof. Given an exact sequence $B \xrightarrow{\beta} C \to 0$, we can always extend it to a short exact sequence
$$0 \to \mathrm{Ker}(\beta) \to B \xrightarrow{\beta} C \to 0.$$
Applying $\mathrm{Hom}_R(M,\cdot)$, the projectivity of M is equivalent to
$$\mathrm{Hom}_R(M,B) \twoheadrightarrow \mathrm{Hom}_R(M,C),$$
which is exactly the content of diagram (32). □

The next lemma provides us with the first examples of projective modules.

Lemma 31.1.2. *A free R-module is projective.*

Proof. Let M be a free R-module; M is isomorphic to $\oplus_{i\in I} R$. To simplify notation, we write
$$M = \oplus_{i\in I} R.$$
Thus, the module M is generated as an R-module by the elements $\{e_i : i \in I\}$, where e_i is the vector with 1 in the i-th coordinate and 0 elsewhere. That means that every element in M has a unique expression as $\sum_{i\in I} r_i e_i$, where only finitely many r_i are non-zero.

Given homomorphisms $f\colon M \to C$ and $\beta\colon B \to C$, where β is surjective as in (32), choose for every $i \in I$ an element $b_i \in B$ such that $\beta(b_i) = f(e_i)$ and define

$$g\colon M \to B, \qquad g(\sum_i r_i e_i) = \sum_i r_i b_i.$$

The sum is well-defined as only finitely many terms are non-zero and g is well-defined due to the uniqueness of expression of elements of M as sums. We have

$$\beta(g(\sum_i r_i e_i)) = \beta(\sum_i r_i b_i) = \sum_i r_i \beta(b_i) = \sum_i r_i f(e_i) = f(\sum_i r_i e_i)$$

and so $\beta \circ g = f$. □

Theorem 31.1.3. *An R-module P is projective if and only if P is isomorphic to a direct summand of a free R-module.*

Proof. Suppose P is projective. Any R-module, in particular P, is a quotient of a free module F. Consider thus a diagram

$$\begin{array}{ccccc} & & & & P \\ & & \overset{g}{\swarrow} & & \downarrow{\scriptstyle id} \\ F & \xrightarrow{\beta} & P & \longrightarrow & 0 \end{array}$$

in which F is free. It follows that $F = \mathrm{Ker}(\beta) \oplus g(P)$ and $P \cong g(P)$. Thus, P is isomorphic to a direct summand of the free module F.

Conversely, suppose that for some module Q, $P \oplus Q$ is a free module. And consider a diagram

$$\begin{array}{ccccc} & & & & P \\ & & \overset{g}{\swarrow} & & \downarrow{\scriptstyle f} \\ B & \xrightarrow{\beta} & C & \longrightarrow & 0. \end{array}$$

We can extend it to a commutative diagram using the natural projection $P \oplus Q \to P$:

$$\begin{array}{ccccc} & & P & \xleftarrow{\pi} & P \oplus Q \\ & & \overset{g}{\swarrow} & & \downarrow{\scriptstyle f} \quad \swarrow{\scriptstyle f_1} \\ B & \xrightarrow{\beta} & C & \longrightarrow & 0. \end{array}$$

Here $f_1(p,q) = f(p)$. As $P \oplus Q$ is free, hence projective by Lemma 31.1.2, there is a map $g_1 : P \oplus Q \to B$ such that $\beta \circ g_1 = f_1$. Now, define $g(p) = g_1(p,0)$.

$$\begin{array}{ccccc} & & P & \xleftarrow{\pi} & P \oplus Q \\ & \overset{g}{\swarrow} & \downarrow{\scriptstyle g_1} & & \downarrow{\scriptstyle f} \quad \swarrow{\scriptstyle f_1} \\ B & \xleftarrow{\beta} & C & \longrightarrow & 0. \end{array}$$

Then, $\beta(g(p)) = \beta(g_1(p,0)) = f_1(p,0) = f(p)$. □

Remark 31.1.4. The module Q such that $P \oplus Q$ is free is also projective, by the same Theorem. Thus, we might say that P is projective if and only if there is a projective module Q such that $P \oplus Q$ is free.

Example 31.1.5. Here are some examples of projective and non-projective modules.

(1) 0, R, R^2, ... are free R-modules hence projective.

(2) If M has torsion as an R-module[46] then M cannot be a submodule of a free module, let alone a direct summand, so M is not projective. For example, $\mathbb{Z}/n\mathbb{Z}$ is not a projective $\mathbb{Z}$-module.
(3) Let K be a number field and I a non-zero ideal of $\mathcal{O}_K$. Then I is a projective $\mathcal{O}_K$-module (though this is not an obvious statement), but need not be free. Such ideals exist if and only if the class group of $\mathcal{O}_K$ is not the trivial group. For example, the ideal $P := \langle 2, \sqrt{-6}\rangle$ of $\mathbb{Z}[\sqrt{-6}]$ is not free. It is projective. In fact, one can prove that $P \oplus P$ is a free rank 2 module over $\mathbb{Z}[\sqrt{-6}]$.
(4) Any direct summand of a projective module is projective.

Proposition 31.1.6. *(1) If P_1, P_2 are projective R-modules so is $P_1 \oplus P_2$.*
(2) If R is a commutative and P_1, P_2 are projective R-modules so is $P_1 \otimes_R P_2$.

Proof. Let Q_1, Q_2 be modules such that $P_1 \oplus Q_1, P_2 \oplus Q_2$ are free R-modules. Then

$$(P_1 \oplus Q_1) \oplus (P_2 \oplus Q_2) \cong (P_1 \oplus P_2) \oplus (Q_1 \oplus Q_2),$$

is free too and $P_1 \oplus P_2$ is a direct summand of it, hence projective. Likewise, we have the free module

$$(P_1 \oplus Q_1) \otimes (P_2 \oplus Q_2) \cong (P_1 \otimes P_2) \oplus (Q_1 \otimes P_2 \oplus P_1 \otimes Q_2 \oplus Q_1 \otimes Q_2)$$

and $P_1 \otimes P_2$ is a direct summand. □

31.2. **The class group.** For simplicity we shall assume throughout this section that R is a commutative ring. Our goal is to define the class group, which is a commutative group whose elements are isomorphism classes of the so-called invertible R-modules. In various situations the class group plays an important role. In algebraic number theory it measures the failure of rings such as the rings of integers of a number field to be unique factorization domains (if they always were, Fermat's Last Theory would have been proven in the 19th century). In algebraic geometry, the class group of a commutative ring A classifies line bundles over an affine scheme $\mathrm{Spec}(A)$. The class group is, in general, an interesting object of study in the theory of commutative rings. As Theorem 31.2.1 shows, the class group measures the local-to-global obstruction for any module to be free.

This topic is skipped this year. Note that the notes on this topic (§31.2) are missing some details and may contain typos.

31.2.1. *Projective modules over local rings.*

Theorem 31.2.1. *Let R be a local ring. Any finitely generated projective module over R is free.*

Proof. Let M be a finitely generated projective R-module. For some integer n, there is a surjection $R^n \to M$ that splits, because M is projective. Thus, there is a finitely generated module N such that $M \oplus N = F$, a free module of finite rank n. We find that as $k := R/J(R)$-modules we have

$$\frac{M}{J(R)M} \oplus \frac{N}{J(R)N} = \frac{F}{J(R)F}.$$

But $J(R)$ is the maximal ideal of R and thus k is a field. That implies that there are finitely many elements $\overline{m_1}, \ldots, \overline{m_a}$ of $\overline{M} := M/J(R)M$ that are a basis for $\overline{M}$ over k. By Corollary 26.2.2, $m_1, \ldots, m_a$ generate M over R.

Complete $\overline{m_1}, \ldots, \overline{m_a}$ by $\overline{n_1}, \ldots, \overline{n_b}$ a basis of $\overline{N} = N/J(R)N$ to get a basis of $\overline{F} = F/J(R)F$. We see that $a + b = n$ and that $m_1, \ldots, m_a, n_1, \ldots, n_b$ is a set of generators for F, which is minimal because

$$\mathrm{rank}_R(F) = \dim_k(\overline{F}) = a + b.$$

Therefore, they forms a basis for F over R. We conclude that $m_1, \ldots, m_a$ are generators for M and independent over R and this implies that M is free. □

[46] Recall that an element m of M is called **torsion** if there is a regular element $r \in R$ such that $rm = 0$. An element r is **regular** element means that for any $s \in R, s \neq 0$, $sr \neq 0$ and $rs \neq 0$. Note that when we consider R as a module over itself, there are no torsion elements.

Recall that for any commutative ring R a free finitely-generated module M is isomorphic to R^n for some integer n that is uniquely determined and called its **rank** (the proof that n is unique can be obtained by reducing modulo a maximal ideal I of R and thereby reducing the statement to vector spaces over R/I).

Let S be a multiplicative set. Suppose that M is a free R-module of rank n then $M[S^{-1}]$ is a free $R[S^{-1}]$-module of rank n. This is quite clear from $R^n[S^{-1}] \cong (R[S^{-1}])^n$. Furthermore, if M is projective over R, then $M[S^{-1}]$ is projective over $R[S^{-1}]$. Indeed, for some Q, $M \oplus Q$ is free and then $(M \oplus Q)[S^{-1}] \cong M[S^{-1}] \oplus Q[S^{-1}]$ is a free $R[S^{-1}]$-module. Also, if M is generated over R by n elements, namely, there is a surjection $R^n \to M$, then by localizing (localizing is an exact functor) we find a surjection $R[S^{-1}]^n \to M[S^{-1}]$, showing that $M[S^{-1}]$ is also generated by n elements (and perhaps by less). In fact, if $m_1, \dots, m_n$ generate M then $\frac{m_1}{1}, \dots, \frac{m_n}{1}$ generate $M[S^{-1}]$.

In particular, if M is projective with n generators, for every prime ideal $\mathfrak{p}$ of R, $M_{\mathfrak{p}}$ is free of some rank $n_{\mathfrak{p}} \leq n$.[47]

Theorem 31.2.2. *Let M be a projective module. The function*

$$\mathrm{Spec}(R) \to \mathbb{Z}, \quad \mathfrak{p} \mapsto n_{\mathfrak{p}},$$

is a continuous function. In particular, if $\mathrm{Spec}(R)$ is connected, then $n_{\mathfrak{p}}$ is constant and we call it the **rank** *of M. We denote it $\mathrm{rk}(M)$.*

Proof. We will need a lemma that is of independent interest. It shows that finitely generated projective modules are locally free in the Zariski topology on $\mathrm{Spec}(R)$.

Lemma 31.2.3. *Let M be a finitely generated projective R-module and $\mathfrak{p} \in \mathrm{Spec}(R)$. There is a basic open set D_s such that $\mathfrak{p} \in D_s$ and M is free over D_s. In algebraic terms: there is an element s of R such that $s \notin \mathfrak{p}$ and such that $M[s^{-1}]$ is a free $R[s^{-1}]$-module.*

Assuming the lemma, let us show that the function $\mathfrak{p} \mapsto n_{\mathfrak{p}}$ is continuous. Since the sets D_f constructed in the lemma cover $\mathrm{Spec}(R)$, it is enough to show that the function $\mathfrak{q} \mapsto n_{\mathfrak{q}}$ is constant on the set D_f. But that follows from the discussion preceding the theorem as $M_{\mathfrak{q}}$ is a localization of $M[f^{-1}]$ (f is but one of the elements in $R \setminus \mathfrak{q}$). It remains to prove the lemma.

We now that $M_{\mathfrak{p}}$ is a free $R_{\mathfrak{p}}$ module of some finite rank $m := n_{\mathfrak{p}}$. Let $\frac{x_1}{s_1}, \dots, \frac{x_m}{s_m}$ be a basis for it. So the $x_i \in M$ and the $s_i \in R \setminus \mathfrak{p}$. Since the elements $s_1, \dots, s_n$ are units in $R[S^{-1}]$, also

$$\frac{x_1}{1}, \dots, \frac{x_m}{1}$$

are a basis. The idea now would be to determine by what element we need to localize so that $\frac{x_1}{1}, \dots, \frac{x_m}{1}$ will form a basis of $M[f^{-1}]$. That is, we have a homomorphism

$$f : R^m \to M, \quad (r_1, \dots, r_m) \mapsto \sum_{i=1}^{m} r_i x_i,$$

which gives an exact sequence

$$0 \longrightarrow \mathrm{Ker}(f) \longrightarrow R^m \xrightarrow{f} M \longrightarrow \mathrm{Coker}(f) \longrightarrow 0\,. \tag{33}$$

Localizing (33) at $\mathfrak{p}$, and using the properties given in the proof of Proposition 3.4.3, we get

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Ker}(f_{\mathfrak{p}}) & \longrightarrow & R_{\mathfrak{p}}^m & \xrightarrow{f} & M_{\mathfrak{p}} & \longrightarrow & \mathrm{Coker}(f_{\mathfrak{p}}) & \longrightarrow & 0\,. \\ & & \| & & & & & & \| & & \\ & & 0 & & & & & & 0 & & \end{array}$$

[47] Typically, an ideal I of the ring of integers $\mathcal{O}_K$ of a number field K is generated by 2 elements and not less, but for every $\mathfrak{p}$ its localization is generated by a single element. That is, $n = 2$ in many situations, but $n_{\mathfrak{p}} = 1$ for all $\mathfrak{p}$ in this case.

Now, $\mathbf{Coker}(f)$, being a homomorphic image of M is finitely generated and $\mathbf{Ker}(f)$ being a sub-module of the finitely generated module R^m (that is noetherian) is also finitely generated. Thus, the generators of those modules become 0 in their localization at $\mathfrak{p}$. But, looking at what this means by definition, it follows that there is an element $s \in R \setminus \mathfrak{p}$ that kills all these generators. That is, also $\mathbf{Ker}(f)[s^{-1}] = 0$ and $\mathbf{Coker}(f)[s^{-1}] = 0$. Localizing (33) at $\{1, s, s^2, \dots\}$, we find that

$$R[s^{-1}]^m \cong M[s^{-1}].$$

□

31.2.2. *Invertible modules.* Let M be a projective module over a finitely generated ring R. We call M an **invertible** module if there is an R-module N such that $M \otimes_R N \cong R$, as R modules.

Lemma 31.2.4. *Let M be a finitely-generated projective module over R. Then M is invertible if and only if $\mathbf{rk}(M) = 1$.*

Proof. Suppose that $\mathbf{rk}(M) = 1$ and $M \otimes_R N \cong R$. Localizing at a prime ideal we find that

$$M_{\mathfrak{p}} \otimes_{R_{\mathfrak{p}}} N_{\mathfrak{p}} \cong R_{\mathfrak{p}}.$$

But M is free finite rank n and thus the left hand side is isomorphic to $N_{\mathfrak{p}}^n$; it follows that $N_{\mathfrak{p}}$ is a direct summand of the free $R_{\mathfrak{p}}$-module $R_{\mathfrak{p}}$, hence projective, hence free, hence isomorphic to $R_{\mathfrak{p}}^m$. It now follows that $m = n = 1$. [48]

Conversely, assume that M is projective of rank 1. We claim that

$$M \otimes_R \mathbf{Hom}(M, R) \cong R.$$

First, there is a natural homomorphism of R-modules

$$M \otimes_R \mathbf{Hom}(M, R) \to R, \quad (m, f) \mapsto f(m).$$

To show this is an isomorphism, it is enough to show that after localizing at every prime ideal $\mathfrak{p}$. But $M_{\mathfrak{p}} \cong R_{\mathfrak{p}}$ and so, modulo checking that localization commutes with $\mathbf{Hom}$ and similar statement,

$$R_{\mathfrak{p}} = \mathbf{Hom}_{R_{\mathfrak{p}}}(M_{\mathfrak{p}}, R_{\mathfrak{p}}) = \mathbf{Hom}(M, R)_{\mathfrak{p}}.$$

Consequently

$$M_{\mathfrak{p}} \otimes_{R_{\mathfrak{p}}} \mathbf{Hom}(M, R)_{\mathfrak{p}} \cong R_{\mathfrak{p}}, \quad \forall \mathfrak{p}.$$

□

31.2.3. *The class group.* Let R be a commutative ring. We define the **class group** $\mathrm{Cl}(R)$ of R to be the isomorphism classes of projective rank 1 modules over R, where the group operation is the tensor product. We proved above that every element is indeed invertible. It should only be mentioned that the tensor product of two projective rank 1 modules is rank 1 and that can be checked by localizing at every prime ideal $\mathfrak{p}$ of R.

The class group is a very important construction. In the context of algebraic geometry, one can show that the projective rank 1 modules over R correspond to line bundles over $\mathbf{Spec}(R)$. Of course, the notion of line bundle must be defined, but it is a notion that is visibly an analogue of the notion of a line bundle over a manifold.

In the context of number theory, the class group is a very important number theoretic invariant. Some of the fundamental results are that if $\mathcal{O}_K$ is the ring of integers of a number field K then $\mathrm{Cl}(\mathcal{O}_K)$ is a finite group that is trivial if and only if $\mathcal{O}_K$ is a unique factorization domain. In fact, one shows that every projective module over $\mathcal{O}_K$ is isomorphic to a fractional ideal of K. That is, it is isomorphic as an $\mathcal{O}_K$-module to a subgroup of K of the form $\frac{1}{n}I$, where $I \lhd \mathcal{O}_K$ is a non-zero ideal. Thus, one can prove that the class group is trivial if and only if $\mathcal{O}_K$ is a principal ideal domain. In particular, the class group can be seen as measuring the

[48] It follows that N is projective because the exactness of the functor $\mathbf{Hom}_R(N, \cdot)$ is a local property.

failure of unique factorization. But we may perhaps mention that if $K=\mathbb{Q}(\sqrt{d})$ where $d>1$ is a square free integer, then $2^e|\sharp\mathrm{Cl}(\mathcal{O}_K)$, where $e+1$ is the number of distinct prime divisors of d. Thus, K can only have class number one in this case if d is prime, but the converse doesn't hold. For example, the class number of $\mathbb{Q}(\sqrt{229})$ is equal to 3. It is an unproven conjecture of Gauss that there are infinitely many primes p such that the class number of $\mathbb{Q}(\sqrt{p})$ is equal to 1.

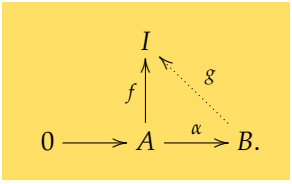
Unfortunately, discussing examples would be too much of a digression. The subject of class groups has been a central area of number theory and algebraic geometry since its inception and remains at their forefront.

32. Injective modules

32.1. **Definition and basic properties.** Recall that an R-module M is called **injective** if $\mathrm{Hom}_R(\cdot,M)$ is exact. The notion of injective has to do with *extending* homomorphisms *into* M.

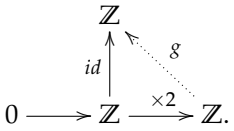
Theorem 32.1.1. *A module I is injective if and only if given a diagram of R-modules with α injective, there is a homomorphism g such that $g\circ\alpha=f$:*

(34)



Proof. We can always complete the exact diagram of R-modules $0\to A\to B$ to a short exact sequence $0\to A\to B\to B/A\to 0$ and then apply $\mathrm{Hom}_R(\cdot,I)$. And, in fact, up to isomorphism any SES looks like. Thus, I is injective, if and only if $\mathrm{Hom}(B,I)\to\mathrm{Hom}(A,I)$ is surjective for any injective map $A\hookrightarrow B$. But this is exactly the content of diagram (34). □

Unlike the situation with projective modules, injective modules are much harder to come by. For example, $\mathbb{Z}$ is not an injective module. One cannot find a homomorphism g such that the following diagram is commutative:

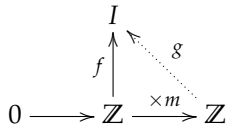


Thus, the question arises whether non-zero injective modules exist at all! We will see that there are plenty of injective modules. The development of the theory begins with the case of $\mathbb{Z}$-modules:

32.2. **Injective $\mathbb{Z}$-modules.** An abelian group I is called **divisible** if for every $a\in I, m\in\mathbb{Z}, m\neq 0$ there is a $b\in I$ such that $mb=a$.

Proposition 32.2.1. *An abelian group I is an injective $\mathbb{Z}$-module if and only if it is divisible.*

Proof. Let I be an injective $\mathbb{Z}$-module. Let $a\in I$. Consider the diagram



where $f(n)=na$. Looking at the images of the element $1\in\mathbb{Z}$ we see that $a=g(m)=mg(1)$. Let $b=g(1)$ then $mb=a$ and we conclude that I is divisble.

Conversely, suppose that I is divisible. Without loss of generality, we may replace an injection by inclusion and so we need to show that following. *Suppose H is a subgroup of an abelian group G and $f\colon H\to I$ is a group homomorphism. Then, there is a group homomorphism $g\colon G\to I$ that extends f.*

To show that, consider the set

$$\{(H', f') : H \leq H', f' \colon H' \to I, f'|_H = f\}$$

of all possible extensions f' (a homomorphism) of f to all possible subgroups H' of G that contain H. There is a natural partial order on this set and chains have an upper bound. Thus, we may apply Zorn's lemma and deduce that there is a maximal element (H_0, f_0). We claim that $H_0 = G$ and thus f_0 is the homomorphism we are after.

Suppose $H_0 \neq G$ and let $g \in G \setminus H_0$. If $\langle g \rangle \cap H_0 = \{0\}$ then $\langle g \rangle + H_0$ is a direct sum and we can extend f_0 (and hence f) to it by setting $f_0(g) = 0$. That's a contradiction. Therefore, for some $N > 0$, we have

$$\langle g \rangle \cap H_0 = \langle Ng \rangle.$$

(And we may assume $Ng \neq 0$, but that doesn't play any role in the following.) Choose an element $a_g \in I$ such that $Na_g = f_0(Ng)$. Define a homomorphism of groups

$$h \colon \langle g \rangle \oplus H_0 \to I, \qquad h(ng, h) = na_g + f_0(h).$$

This homomorphism is well defined. If $ng = mg$ then we need to show that $(n-m)a_g = 0$. That is, we need to show that if $dg = 0$ then $da_g = 0$. But, if $dg = 0$ then $dg \in \langle g \rangle \cap H_0$ and so $d = kN$ for some integer k. Then, on the one hand since $dg = 0$ we have $f_0(dg) = 0$, but on the other hand, $f_0(dg) = f_0(kNg) = kf_0(Ng) = kNa_g = da_g$ and it follows that $da_g = 0$.

There is a surjection $\langle g \rangle \oplus H_0 \to \langle g \rangle + H_0 \subseteq G$ with kernel $\mathbb{Z} \cdot (Ng, -Ng)$ and h vanishes on the kernel. Thus, h induces a well-defined homomorphism

$$h_0 \colon \langle g \rangle + H_0 \to I, \quad h_0(ng + h) = na_g + f(h).$$

The pair $(\langle g \rangle + H_0, h_0)$ extends (H_0, f_0) and we get a contradiction again. □

Remark 32.2.2. One can define the notion of being divisible for any ring R, in fact. The relation to being injective is not easily stated. However, one can prove, by modifying the proof above, that if R is a commutative PID then an R-module is injective if and only if it is divisible.

With this Proposition we see that there are plenty of injective $\mathbb{Z}$-modules. For example:

Corollary 32.2.3. *$\mathbb{Q}$ and $\mathbb{Q}/\mathbb{Z}$ are injective $\mathbb{Z}$-modules.*

32.3. **Injective R-modules.** The following theorem shows that injective modules exist in abundance. It is a crucial first step in homological algebra constructions as we shall see later.

Theorem 32.3.1. *Let R be a ring and let M be an R-module. There is an injective R-module I and an injective homomorphism $M \hookrightarrow I$.*

Proof. The first step is to prove the theorem for abelian groups. The proof of the following Lemma is straighforward.

Lemma 32.3.2. *(1) A direct sum $\oplus_i A_i$ of divisible abelian groups is divisible.*
(2) A quotient of a divisible abelian group is divisible.

Let A be an abelian group and write A as a quotient on the free abelian group $\oplus_{a \in A} \mathbb{Z} \cdot [a]$ (here $[a]$ is a formal symbol associated to a, as we do for group rings in general). We can think about its elements as vectors $\sum_{a \in A} n_a \cdot [a]$ with $n_a \in \mathbb{Z}$ and only finitely many of them not zero. There is a natural surjection taking $\sum_{a \in A} n_a \cdot [a]$ to $\sum_{a \in A} n_a a$. And we conclude that we have an isomorphism

$$A \cong (\oplus_{a \in A} \mathbb{Z} \cdot [a])/S,$$

for some subgroup S.

Now, the inclusion $\oplus_{a\in A}\mathbb{Z}\cdot[a]\hookrightarrow\oplus_{a\in A}\mathbb{Q}\cdot[a]$ induces an injection

$$(\oplus_{a\in A}\mathbb{Z}\cdot[a])/S\hookrightarrow(\oplus_{a\in A}\mathbb{Q}\cdot[a])/S$$

and by Lemma 32.3.2 the right hand side is divisible, hence an injective $\mathbb{Z}$-module.

We are now consider the general statement. Let R be a ring and $M\in{}_R\mathbf{Mod}$. Considering M as an abelian group, we have an inclusion

$$M\hookrightarrow D,$$

where D is a divisible abelian group. Consider $\mathrm{Hom}_{\mathbb{Z}}(R,D)$. One check directly that this abelian group is an R-module, where for $r\in R, f\in\mathrm{Hom}_{\mathbb{Z}}(R,D)$ we define $rf\colon R\to D$ by

$$(rf)(s)=f(sr).$$

It is easy to verify that $(rf)(s_1+s_2)=(rf)(s_1)+(rf)(s_2)$ so $rf\in\mathrm{Hom}_{\mathbb{Z}}(R,D)$ and $(r(f_1+f_2))(s)=(rf_1+rf_2)(s)$, $((r_1r_2)f)(s)=f(sr_1r_2)=(r_2f)(sr_1)=(r_1(r_2f))(s)$, etc.; it follows that we defined a module structure on $\mathrm{Hom}_{\mathbb{Z}}(R,D)$.

We have an embedding of R-modules

$$M\hookrightarrow\mathrm{Hom}_{\mathbb{Z}}(R,D),\quad m\mapsto\varphi_m,$$

where $\varphi_m(r)=rm$. Clearly, $\varphi_{m_1+m_2}=\varphi_{m_1}+\varphi_{m_2}$, and $\varphi_{rm}=r\varphi_m$, because $\varphi_{rm}(s)=srm=\varphi_m(sr)=(r\varphi_m)(s)$. The injectivity follows once one notes that $\varphi_m(1)=m$. Thus, it remains to show that $\mathrm{Hom}_{\mathbb{Z}}(R,D)$ is an injective R-module. To clarify the argument we state this as a separate proposition.

Proposition 32.3.3. *Let D be a divisible abelian group then $\mathrm{Hom}_{\mathbb{Z}}(R,D)$ is an injective left R-module, where $(rf)(x)=f(xr)$.*

Proof. (Proposition) We show that the functor $\mathrm{Hom}_R(-,\mathrm{Hom}_{\mathbb{Z}}(R,D))$ is exact. It is enough to show that for an exact sequence in ${}_R\mathbf{Mod}$, $0\to A\to B$, the arrow at the top row of the following diagram is surjective. We indicate it by putting a double question mark about the double-head.

$$\begin{array}{ccc}
\mathrm{Hom}_R(B,\mathrm{Hom}_{\mathbb{Z}}(R,D)) & \overset{??}{\twoheadrightarrow} & \mathrm{Hom}_R(A,\mathrm{Hom}_{\mathbb{Z}}(R,D)) \\
\downarrow\cong & & \downarrow\cong \\
\mathrm{Hom}_{\mathbb{Z}}(R\otimes_R B,D) & \overset{??}{\twoheadrightarrow} & \mathrm{Hom}_{\mathbb{Z}}(R\otimes_R A,D) \\
\downarrow\cong & & \downarrow\cong \\
\mathrm{Hom}_{\mathbb{Z}}(B,D) & \overset{??}{\twoheadrightarrow} & \mathrm{Hom}_{\mathbb{Z}}(A,D)
\end{array}$$

The diagram is easily verified to be commutative; we have used the adjoint property of $\otimes-\mathrm{Hom}$ for the first vertical isomorphism. The surjectivity of the lowest horizontal arrow would imply that of the top arrow. Because D is a divisible abelian group, hence an injective abelian group, the exact sequence $0\to A\to B$ (now viewed as a sequence of abelian groups) gives an exact sequence $\mathrm{Hom}_{\mathbb{Z}}(B,D)\to\mathrm{Hom}_{\mathbb{Z}}(A,D)\to 0$ so the lower row is indeed surjective. □

□

The theorem gives us a method to constructive injective modules.

Corollary 32.3.4. *For any ring R, $\mathrm{Hom}_{\mathbb{Z}}(R,\mathbb{Q})$ and $\mathrm{Hom}_{\mathbb{Z}}(R,\mathbb{Q}/\mathbb{Z})$ are injective R-modules.*

Finally, we mention without proof a very nice characterization of injective modules.

Theorem 32.3.5 (Baer's Criterion)**.** *Let M be an R-module. Then M is injective if and only if for every ideal I of R, every R-homomorphism $I \to M$ can be extended to R.*

33. Flat modules

Let R be a ring. A module $M \in {}_{\mathbf{R}}\mathbf{Mod}$ is called **flat** if the functor $(-) \otimes_R M$ is an exact functor $\mathbf{Mod_R} \to \mathbf{AbGps}$. Since we have proved that for an exact sequence in $\mathbf{Mod_R}$

$$0 \to A_1 \to A_2 \to A_3 \to 0$$

we get an exact sequence

$$A_1 \otimes_R M \to A_2 \otimes_R M \to A_3 \otimes_R M \to 0,$$

we see that M if flat if and only if

$$A_1 \hookrightarrow A_2 \implies A_1 \otimes_R M \hookrightarrow A_2 \otimes_R M.$$

That is, M is flat if and only tensoring with M preserves injectivity of maps.

33.1. **Some basic examples of flat and non-flat modules.**

Proposition 33.1.1. *A projective module is flat.*

Proof. If F is a free R-module, $F \cong \oplus_{i\in I} R$, then F is flat because

$$A_1 \otimes_R F \cong A_1 \otimes_R (\oplus_{i\in I} R) \cong \oplus_{i\in I} A_1 \otimes_R R \cong \oplus_{i\in I} A_1 \hookrightarrow \oplus_{i\in I} A_2 \cong A_2 \otimes_R F.$$

(The last isomorphism is obtained by the same argument used for A_1.)

If P is projective, then for some R-module Q, $P \oplus Q = F$, a free R-module. Then

$$A_1 \otimes_R (P \oplus Q) \hookrightarrow A_2 \otimes_R (P \oplus Q).$$

But $A_i \otimes_R (P \oplus Q) = A_i \otimes_R P \oplus A_i \otimes_R Q$ and thus $A_1 \otimes_R P \hookrightarrow A_2 \otimes_R P$. □

Example 33.1.2. $\mathbb{Z}/2\mathbb{Z}$ is not a flat $\mathbb{Z}$-module, because tensoring $0 \to \mathbb{Z} \to \mathbb{Q}$ with $\mathbb{Z}/2\mathbb{Z}$ we get $0 \to \mathbb{Z}/2\mathbb{Z} \to 0$, which is not exact. This example is predicted by the following theorem (that we will not prove):

Theorem 33.1.3. *If M is a* **finitely presented** *R-module*[49] *then M is flat if and only if M is projective.*

Theorem 33.1.4. *Let R be a commutative ring and S a multiplicative set in R then $R[S^{-1}]$ is a flat R-module.*

Proof. Indeed, we proved that localization is an exact functor (Proposition 3.2.3). □

So for example, if R is an integral domain with field of fractions Q then Q is a flat R-module. In particular, $\mathbb{Q}$ is a flat $\mathbb{Z}$-module. Note that $\mathbb{Q}$ is not projective because it does no embed into a free $\mathbb{Z}$-module (no non-zero free $\mathbb{Z}$-module is divisible). Thus, Theorem 33.1.3 may fail for non finitely presented modules.

[49] That means that there is an exact sequence $F_1 \to F_2 \to M \to 0$ of R-modules where F_i are free R-modules of finite length. We can think about that as M being finitely generated due to the surjection $F_2 \twoheadrightarrow M$ and the module of relations between the generators, namely the image of $F_1 \to F_2$, being finitely generated as well.

33.2. **Relations between flat, projective and injective.** We have the following statements connecting the notions of projective, injective and flat:

(1) Projective $\Longrightarrow$ Flat (Proposition 33.1.1).
(2) Flat and finitely presented $\Longrightarrow$ Projective (Theorem 33.1.3).
(3) Projective $\not\Longrightarrow$ Injective ($\mathbb{Z}$).
(4) Injective $\not\Longrightarrow$ Projective ($\mathbb{Q}/\mathbb{Z}$, $\mathbb{Q}$).
(5) Injective $\not\Longrightarrow$ Flat ($\mathbb{Q}/\mathbb{Z}$).
(6) Flat $\not\Longrightarrow$ Injective ($\mathbb{Z}$).
(7) Flat $\not\Longrightarrow$ Projective ($\mathbb{Q}$).

On the positive side we have the following. Let $B \in \mathbf{Mod_R}$ and define its **character module** B^* to be the left R-module

$$B^* = \mathrm{Hom}_{\mathbb{Z}}(B, \mathbb{Q}/\mathbb{Z}).$$

One can prove that $B \to B^*$ is an exact functor $\mathbf{Mod_R} \to {}_{\mathbf{R}}\mathbf{Mod}$ and that

$$B \in \mathbf{Mod_R} \text{ is flat } \iff B^* \in {}_{\mathbf{R}}\mathbf{Mod} \text{ is injective.}$$

A common application of flatness is given by the following proposition. It is especially useful when R is a commutative ring because in that case both $I \otimes_R M$ and IM are R-modules.

Proposition 33.2.1. *Let M be an R-module. If M is a flat R-module and I is a right R-ideal then the natural map*

$$I \otimes_R M \to IM$$

is an isomorphism.

Proof. By IM we mean the subgroup M of element of the form $\sum i_n m_n$, $i_n \in I, m_n \in M$. It is an R-module if I is a two-sided ideal, but not in general.

Assume M is flat. Then, as $I \hookrightarrow R$ as R-modules, $I \otimes_R M \hookrightarrow R \otimes_R M \cong M$ and the image is precisely IM. □

Remark 33.2.2. The converse is true too. If for every such ideal I the natural map $I \otimes_R M \to IM$ is an isomorphism then M is flat. The proof requires use of the character module and we omit it.

34. Complexes and homology

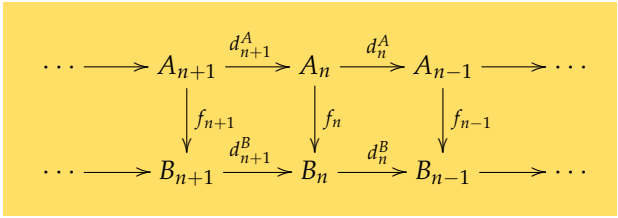
34.1. **Complexes.** Let R be a ring. An R-**complex** $(A_\bullet, d^A_\bullet)$ is a diagram of R-modules and R homomorphisms

$$\cdots \longrightarrow A_{n+1} \xrightarrow{d^A_{n+1}} A_n \xrightarrow{d^A_n} A_{n-1} \longrightarrow \cdots$$

indexed by $\mathbb{Z}$ such that for all n, $d^A_n \circ d^A_{n+1} = 0$. The maps d^A_n are called **differentials**. A **morphism of complexes**

$$f_\bullet \colon (A_\bullet, d^A_\bullet) \to (B_\bullet, d^B_\bullet)$$

is a collection of R-module homomorphisms $f_n \colon A_n \to B_n$ such that all the squares in the following diagram commute:



R-complexes form a category $R-\mathbf{Comp}$. This category is abelian. In particular:

- $\mathrm{Ker}(f_\bullet) = (\{\mathrm{Ker}(f_n)\}, \{d^A_n\})$ is a complex.
- $\mathrm{Im}(f_\bullet) = (\{\mathrm{Im}(f_n)\}, \{d^B_n\})$ is a complex.
- $\mathrm{Coker}(f_\bullet) = (\{\mathrm{Coker}(f_n)\}, \{d^B_n\})$ is a complex.

Normally, there will be no need to decorate the differentials d^A_n with so many indices and we will simply write d. The complex condition is then written as $d^2 = 0$.

Example 34.1.1. Let R and S be two rings and $F\colon R-\mathbf{Mod} \to S-\mathbf{Mod}$ be a covariant additive functor. (We write $R-\mathbf{Mod}$ so as not to commit to either left or right modules, but we are either talking about left R-modules and left S-modules, or about right R-modules and right S modules.) Then F extends to an additive functor

$$F\colon R-\mathbf{Comp} \to S-\mathbf{Comp}, \qquad F(A_\bullet, d_\bullet) = (\{FA_n\}, \{Fd_n\}).$$

The later is indeed a complex because $F(d_n) \circ F(d_{n+1}) = F(d_n \circ d_{n+1}) = F(0) = 0$, since F is additive.

Remark 34.1.2. This also works for a contravariant additive functor if one reindexes the objects $F(A_n)$, say by letting $C_{-n} := F(A_n)$, so that the differentials still reduce the degrees.

34.2. **The homology of a complex.** Let $(A_\bullet, d_\bullet)$ be a complex of R-modules. We define its homology $h_\bullet(A_\bullet)$[50] by

$$h_n(A_\bullet) := \frac{\mathrm{Ker}(d_n\colon A_n \to A_{n-1})}{\mathrm{Im}(d_{n+1}\colon A_{n+1} \to A_n)}$$

Thus, $h_\bullet(A_\bullet)$ is a collection of R-modules indexed by $\mathbb{Z}$. We make this into an R-complex by decreeing all differentials to be 0. Then we can state:

Lemma 34.2.1. *$h_\bullet$ is a covariant additive functor*

$$h_\bullet\colon R-\mathbf{Comp} \;\to\; R-\mathbf{Comp}.$$

Proof. The content of this lemma is that given a morphism $f_\bullet\colon A_\bullet \to B_\bullet$ there is a naturally induced homomorphism

$$\frac{\mathrm{Ker}(d_n\colon A_n \to A_{n-1})}{\mathrm{Im}(d_{n+1}\colon A_{n+1} \to A_n)} \xrightarrow{h_n(f)} \frac{\mathrm{Ker}(d_n\colon B_n \to B_{n-1})}{\mathrm{Im}(d_{n+1}\colon B_{n+1} \to B_n)}.$$

This follows from the fact that $f_\bullet$ commutes with the differentials and so f_n takes $\mathrm{Ker}(d^A_n)$ to $\mathrm{Ker}(d^B_n)$ and $\mathrm{Im}(d^A_{n+1})$ to $\mathrm{Im}(d^B_{n+1})$. The other requirements, such as $h_\bullet(f_1 + f_2) = h_\bullet(f_1) + h_\bullet(f_2)$, $h_\bullet((A_\bullet \oplus B_\bullet, d^A_\bullet \oplus d^B_\bullet)) \cong h_\bullet(A_\bullet, d^A_\bullet) \oplus h_\bullet(B_\bullet, d^B_\bullet)$ and $h_\bullet(0_\bullet) = 0_\bullet$, are clear from the definitions. □

[50] In order to simplify the notation, we drop the differentials from the notation $h_\bullet(A_\bullet)$.

Remark 34.2.2. The Lemma doesn't contain more information than saying that for every n

$$h_n\colon R\text{−}\mathbf{Comp} \to R\text{−}\mathbf{Mod}$$

is a covariant additive functor.

Example 34.2.3. (1) If $(A_\bullet, d_\bullet)$ is exact, namely if for every n we have $\mathrm{Im}(d_{n+1}) = \mathrm{Ker}(d_n)$, then $h_n(A_\bullet) = 0$ for every n.

(2) If all the differentials $d_n = 0$ then $h_\bullet(A_\bullet) = A_\bullet$.

(3) If $(A_\bullet, d_\bullet)$ is exact and $F\colon R\text{−}\mathbf{Mod} \to S\text{−}\mathbf{Mod}$ is an additive functor then $F(A_\bullet)$ need not be exact and, somehow, *this is the whole point of homological algebra*. For example, suppose that R is a commutative ring and M is an R-module. We have an additive right-exact functor

$$(\cdot) \otimes_R M\colon R\text{−}\mathbf{Mod} \to R\text{−}\mathbf{Mod},$$

which in general is not exact. Given a short exact sequence $0 \to E \to F \to G \to 0$ of R-modules (thought of as a complex when we extend it in each direction by 0's) we only have the exactness of $M \otimes E \to M \otimes F \to M \otimes G \to 0$, but the first arrow need not be injective.

The machinery of homological algebra provides us with canonical modules $\mathrm{Tor}_n(M,E)$, $\mathrm{Tor}_n(M,F)$, $\mathrm{Tor}_n(M,G)$, and homomorphisms, so that the following sequence is exact:

$$\cdots \to \mathrm{Tor}_2(M,G) \to \mathrm{Tor}_1(M,E) \to \mathrm{Tor}_1(M,F) \to \mathrm{Tor}_1(M,G) \to M \otimes E \to M \otimes F \to M \otimes G \to 0\,.$$

34.3. The Snake Lemma and exactness of $h_\bullet$.

Lemma 34.3.1 (The Snake Lemma)**.** *Let*

$$\begin{array}{ccccccc} A_1 & \xrightarrow{\alpha_1} & B_1 & \xrightarrow{\beta_1} & C_1 & \longrightarrow & 0 \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & A_2 & \xrightarrow{\alpha_2} & B_2 & \xrightarrow{\beta_2} & C_2 \end{array}$$

be a commutative diagram of R-modules with exact rows. Then the sequence

$$\begin{array}{ccccccc} \mathrm{Ker}(f) & \xrightarrow{\alpha_1} & \mathrm{Ker}(g) & \xrightarrow{\beta_1} & \mathrm{Ker}(h) & \xrightarrow{\delta} & \\ \mathrm{Coker}(f) & \xrightarrow{\alpha_2} & \mathrm{Coker}(g) & \xrightarrow{\beta_2} & \mathrm{Coker}(h) & & \end{array}$$

is exact, where δ (the **connecting homomorphism***) is defined as follows: given an element c_1 in $\mathrm{Ker}(h)$ lift it to an element $b_1 \in B$; note that $\beta_2(g(b_1)) = h(\beta_1(b_1)) = h(c_1) = 0$. Thus, there is an $a_2 \in A_2$ such that $\alpha_2(a_2) = g(b_1)$. Define,*

$$\delta(c_1) = a_2.$$

Moreover, if we start with a diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \xrightarrow{\alpha_1} & B_1 & \xrightarrow{\beta_1} & C_1 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & A_2 & \xrightarrow{\alpha_2} & B_2 & \xrightarrow{\beta_2} & C_2 & \longrightarrow & 0 \end{array}$$

with exact rows, then the following is an exact sequence:

$$0 \longrightarrow \mathrm{Ker}(f) \xrightarrow{\alpha_1} \mathrm{Ker}(g) \xrightarrow{\beta_1} \mathrm{Ker}(h) \xrightarrow{\delta} \mathrm{Coker}(f) \xrightarrow{\alpha_2} \mathrm{Coker}(g) \xrightarrow{\beta_2} \mathrm{Coker}(h) \longrightarrow 0.$$

Proof. The proof is left as an exercise. It is mostly straight-forward, but requires a lot of patience. The most delicate points are to show that δ is well-defined and that it is exact at its origin and its target. We show that δ is well-defined and exact at its target.

Given $c_1 \in \mathrm{Ker}(h)$ suppose we choose another element b_1' such that $\beta_1(b_1') = c_1$. We get the element $g(b_1')$ which is of the form $\alpha_2(a_2')$. We have $\alpha_2(a_2 - a_2') = g(b_1 - b_1')$. Now $\beta_1(b_1 - b_1') = 0$ so $b_1 - b_1' = \alpha_1(a_1)$ for some $a_1 \in A_1$. Therefore, $g(b_1 - b_1') = \alpha_2(f(a_1))$, and it follows that $\alpha_2(a_2 - a_2' - f(a_1)) = 0$. As α_2 is injective, $a_2 = a_2' + f(a_1)$, which means that $a_2 = a_2'$ in $\mathrm{Coker}(f)$.

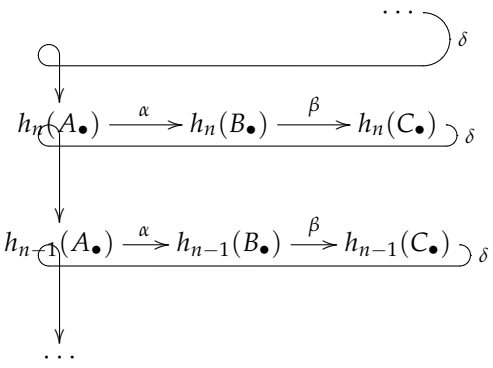
It is easy to check that δ is a homomorphism of R-modules (as all choices can be "coordinated" with addition and multiplication by a scalar). We next check that $\alpha_2 \colon \mathrm{Coker}(f) \to \mathrm{Coker}(g)$ vanishes on the image of δ. Indeed, in the notation above $\alpha_2(a_2 + f(A_1)) = g(b_1) + g(B_1) = g(B_1)$ which means it's the zero element of $\mathrm{Coker}(g)$.

Conversely, suppose that for some element $a_2 + f(A_1) \in \mathrm{Coker}(f)$ we have $\alpha_2(a_2 + f(A_1)) = \alpha_2(a_2) + g(B_1)$ is zero in $\mathrm{Coker}(g)$. This means that $\alpha_2(a_2) = g(b_1)$ for some $b_1 \in B_1$. Let $c_1 = \beta_1(b_1)$. Note that $h(c_1) = h(\beta_1(b_1)) = \beta_2(g(b_1)) = \beta_2(\alpha_2(a_2)) = 0$. So $c_1 \in \mathrm{Ker}(h)$ and we can calculate $\delta(c_1)$. We can choose the same b_1 for the calculation. Then the element $a_1 = \delta(c_1)$ has the property that $\alpha_2(a_1) = g(b_1) = \alpha_2(a_2)$ and so $a_1 = a_2 \in \mathrm{Im}(\delta)$. □

Theorem 34.3.2 (Long exact homology sequence)**.** *An exact sequence of complexes*

$$0 \longrightarrow A_\bullet \xrightarrow{\alpha_\bullet} B_\bullet \xrightarrow{\beta_\bullet} C_\bullet \longrightarrow 0$$

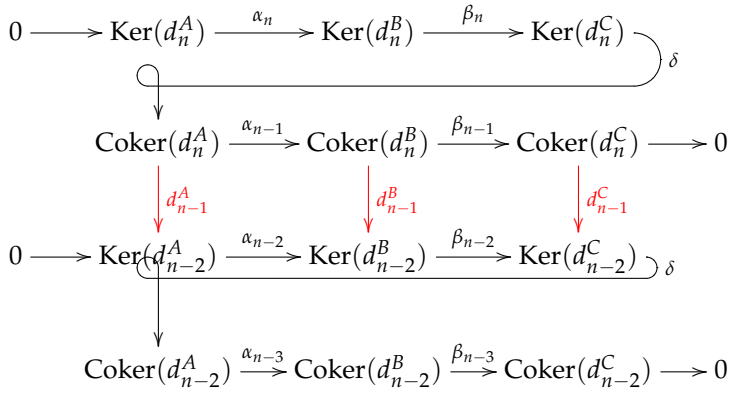
induces a long exact sequence in homology:



Proof. This is an application of the Snake Lemma. For every n we have

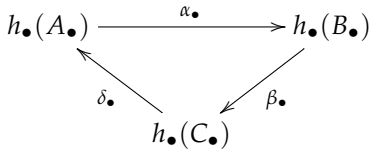
$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_n & \xrightarrow{\alpha_n} & B_n & \xrightarrow{\beta_n} & C_n & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle d_n} & & \downarrow{\scriptstyle d_n} & & \downarrow{\scriptstyle d_n} & & \\ 0 & \longrightarrow & A_{n-1} & \xrightarrow{\alpha_{n-1}} & B_{n-1} & \xrightarrow{\beta_{n-1}} & C_{n-1} & \longrightarrow & 0 \end{array}$$

The Snake Lemma gives the pair of the two upper rows and also the pair of the two lower rows in the following diagram, while the red arrows are naturally induced.

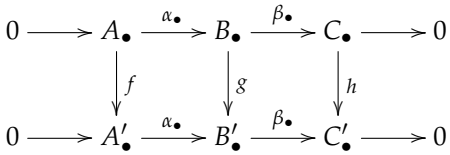


Applying the snake lemma to the two middle rows, we now get the theorem for the $n-1$ and $n-2$ rows, but this hold for every n and the theorem follows. □

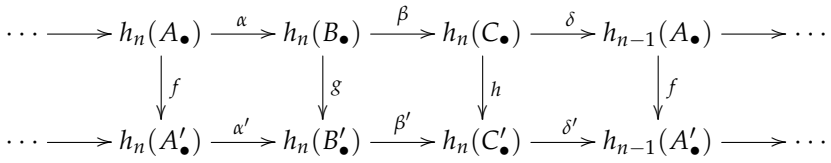
Remark 34.3.3. One can also phrase the theorem by saying that there is an exact triangle of complexes, where $\delta_\bullet$ lowers degrees by -1:



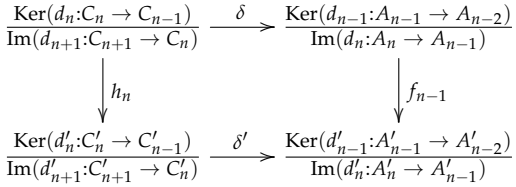
Lemma 34.3.4 (Naturality of the triangle)**.** *A commutative diagram of complexes with exact rows*



induces a commutative diagram on the homology long sequences:



Proof. We already know that the rows are exact and that there are natural maps $f_\bullet, g_\bullet, h_\bullet$. Thus, the new content is that the connecting homomorphisms δ, δ' are also natural. The proof boils down to the statement that the following diagram commutes for every n:



This is actually rather easy to check. Given $c_n \in \mathrm{Ker}(d_n\colon C_n \to C_{n-1})$ we have $c'_n := h_n(c_n) \in \mathrm{Ker}(d'_n\colon C'_n \to C'_{n-1})$ and choosing a $b_n \in B_n$ to calculate $\delta(c_n)$ (so $\beta_n(b_n) = c_n$), we can take $b'_n := g_n(b_n) \in B'_n$ to calculate $\delta'(c'_n)$ because $\beta'_n(b'_n) = \beta'_n(g_n(b_n)) = h_n(\beta_n(b_n)) = h_n(c_n) = c'_n$. The next step, which is that $d_n(b_n) = \alpha_{n-1}(a_{n-1})$ for some $a_{n-1} \in \mathrm{Ker}(d_{n-1}\colon A_{n-1} \to A_{n-2})$ gives, for $a'_{n-1} = f_{n-1}(a_{n-1})$, which is in the kernel of d'_{n-1}, that

$$\alpha'_{n-1}(a'_{n-1}) = \alpha'_{n-1}(f_{n-1}(a_{n-1})) = g_{n-1}(\alpha_{n-1}(a_{n-1})) = g_{n-1}(d_n(b_n)) = d'_n(g_n(b_n)) = d'_n(b'_n),$$

and so $\delta'(c'_n) = a'_{n-1}$. As $\delta(c_n) = a_{n-1}$, this shows that $\delta'(h_n(c_n)) = f_{n-1}(\delta(c_n))$. □

35. Derived functors – I

35.1. **Enough injectives and enough projectives.** Let $\mathbf{A}$ be an abelian category. We say that $\mathbf{A}$ has:

- **enough injectives** if for every object M of $\mathbf{A}$ there is an exact sequence
$$0 \longrightarrow M \longrightarrow I_0 \longrightarrow I_1 \longrightarrow I_2 \longrightarrow \cdots,$$
where I_j is injective for all $j \geq 0$. This is called an **injective resolution** of A and is sometimes denoted
$$0 \to A \to I_\bullet.$$
- **enough projectives** if for every object M of $\mathbf{A}$ there is an exact sequence
$$\cdots \longrightarrow P_2 \longrightarrow P_1 \longrightarrow P_0 \longrightarrow M \longrightarrow 0,$$
where P_j is projective for all $j \geq 0$. This is called a **projective resolution** of A and is sometimes denoted
$$P_\bullet \to A \to 0.$$

Proposition 35.1.1. *The category of R-modules has enough injectives and enough projectives.*

Proof. We construct an injective resolution for M by induction. By Theorem 32.3.1, there is an injective module I_0 and an exact sequence $0 \to M \to I_0$. Suppose we have already shown for some $n \neq 0$ that there is an injective resolution

$$0 \to I_0 \to I_1 \to \ldots \to I_{n-1} \xrightarrow{\alpha_{n-1}} I_n.$$

There is an injection $I_n/\alpha_{n-1}(I_{n-1}) \hookrightarrow I_{n+1}$ for some injective module I_{n+1}. Letting α_n be the composition $I_n \to I_n/\alpha_{n-1}(I_{n-1}) \hookrightarrow I_{n+1}$, we find an exact sequence

$$0 \to I_0 \to I_1 \to \ldots \to I_{n-1} \xrightarrow{\alpha_{n-1}} I_n \xrightarrow{\alpha_n} I_{n+1}.$$

Similarly, we construct a projective resolution (in fact, a **free resolution**) inductively. We can find a free R-module F_0 and a surjection $F_0 \twoheadrightarrow M$. Arguing inductively, suppose that we already found a resolution by free modules

$$F_n \xrightarrow{\alpha_n} F_{n-1} \to \ldots \to F_1 \to F_0 \to M \to 0.$$

There is a surjection $\alpha_{n+1}\colon F_{n+1} \twoheadrightarrow \mathrm{Ker}(\alpha_n)$ from some free module F_{n+1}, which we view as a homomorphism $\alpha_{n+1}\colon F_{n+1} \to F_n$. We then have an exact sequence

$$F_{n+1} \xrightarrow{\alpha_{n+1}} F_n \xrightarrow{\alpha_n} F_{n-1} \to \ldots \to F_1 \to F_0 \to M \to 0.$$

As free modules are projective (Lemma 31.1.2), we found a projective resolution. □

Remark 35.1.2. If $(X, \mathcal{O}_X)$ is a locally ringed space then the category of sheaves of $\mathcal{O}_X$-modules has enough injectives. This is absolutely crucial to geometry (algebraic, complex, differential and so-on). Apparently, it does not have enough projectives.

35.2. **The derived functors of a covariant right-exact functor.** The following definition is one of the key constructions of homological algebra. Although it can be made for any abelian category with enough projectives, we state it for categories of modules as this is our main interest in this course.

Let $F\colon R-\mathbf{Mod} \to S-\mathbf{Mod}$ be a covariant right-exact functor, meaning

$$0 \to A \to B \to C \to 0 \text{ exact} \implies FA \to FB \to FC \to 0 \text{ exact}.$$

We define its **left derived functors**

$$L_nF, \quad n \geq 0.$$

Key Example: The functor $(\cdot) \otimes_R M$ from the category of right R-modules to the category of abelian groups (or even R-modules if R is commutative) is covariant right-exact. The same for $M \otimes_R (\cdot)$.

We define L_nF as follows. Choose a projective resolution $P_\bullet \to A \to 0$. Then $F(P_\bullet)$ is a complex and we take its homology: $L_nF(A) = h_n(FP_\bullet)$. More precisely, we move from the exact complex

$$\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \xrightarrow{\epsilon} A \longrightarrow 0$$

to the complex

$$\cdots \xrightarrow{d_3} P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \longrightarrow 0$$

(that usually will not be exact at P_0). We apply F to get a complex

$$\cdots \xrightarrow{Fd_3} FP_2 \xrightarrow{Fd_2} FP_1 \xrightarrow{Fd_1} FP_0 \longrightarrow 0,$$

and we let

$$L_nF(A) = \frac{\mathrm{Ker}(FP_n \to FP_{n-1})}{\mathrm{Im}(FP_{n+1} \to FP_n)} = \frac{\mathrm{Ker}(Fd_n)}{\mathrm{Im}(Fd_{n+1})}, \quad n \geq 1, \qquad (L_0F)(A) = FP_0/\mathrm{Im}(Fd_1).$$

Remark 35.2.1. As F is right-exact, the sequence

$$FP_1 \xrightarrow{Fd_1} FP_0 \xrightarrow{F\epsilon} FA \longrightarrow 0,$$

is *exact*. Thus, $L_0F(A) = FP_0/\mathrm{Im}(Fd_1) \cong FA$. Thus, as functors,

$$L_0F \cong F$$

Notation: If $F = (\cdot) \otimes_R M$, one denotes

$$\mathrm{Tor}_n(\cdot, M) := L_nF(\cdot).$$

Our goal is to show that L_nF are independent of the choice of resolution and are additive functors. We start with independence on resolutions. We will require the following lemma, which is key to much that follows.

Lemma 35.2.2 (Comparison Lemma)**.** *Consider a commutative diagram*

$$\begin{array}{ccccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\
 & & & & & & & & \downarrow{\scriptstyle f} & & \\
\cdots & \xrightarrow{d'_3} & Q_2 & \xrightarrow{d_2} & Q_1 & \xrightarrow{d'_1} & Q_0 & \xrightarrow{\epsilon'} & B & \longrightarrow & 0
\end{array}$$

where both rows are complexes, the P_i are projective modules and the bottom row is exact. Then f extends to a morphism of complexes

$$\begin{array}{ccccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle f} & & \\
\cdots & \xrightarrow{d'_3} & Q_2 & \xrightarrow{d'_2} & Q_1 & \xrightarrow{d'_1} & Q_0 & \xrightarrow{\epsilon'} & B & \longrightarrow & 0
\end{array} \tag{35}$$

and any two such extensions are **homotopic** *after deleting A and B.*

Remark 35.2.3. The meaning of **homotopic** is the following. Suppose we have two extensions $f_\bullet$ and $h_\bullet$ of f,

$$\begin{array}{ccccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\
 & & {\scriptstyle f_2}\downarrow\downarrow{\scriptstyle h_2} & & {\scriptstyle f_1}\downarrow\downarrow{\scriptstyle h_1} & & {\scriptstyle f_0}\downarrow\downarrow{\scriptstyle h_0} & & \downarrow{\scriptstyle f} & & \\
\cdots & \xrightarrow{d'_3} & Q_2 & \xrightarrow{d'_2} & Q_1 & \xrightarrow{d'_1} & Q_0 & \xrightarrow{\epsilon'} & B & \longrightarrow & 0
\end{array}$$

Let $g_n = h_n - f_n$. Then we require the existence of homomorphisms s_i as in the diagram

$$\begin{array}{ccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & 0 \\
 & \swarrow^{s_2} & \downarrow{\scriptstyle g_2} & \swarrow^{s_1} & \downarrow{\scriptstyle g_1} & \swarrow^{s_0} & \downarrow{\scriptstyle g_0} & \swarrow^{s_{-1}} & \\
\cdots & \xrightarrow{d'_3} & Q_2 & \xrightarrow{d'_2} & Q_1 & \xrightarrow{d'_1} & Q_0 & \xrightarrow{\epsilon'} & 0
\end{array}$$

such that we have the relations

$$g_n = s_{n-1}d_n + d'_{n+1}s_n, \quad n \geq 0.$$

The point of this is that g_n induces the zero map

$$h_\bullet(P_\bullet) \to h_\bullet(Q_\bullet).$$

Indeed, if $x \in \mathrm{Ker}(d_n \colon P_n \to P_{n-1})$ then

$$g_n(x) = s_{n-1}(d_n(x)) + d'_{n+1}(s_n(x)) = d'_{n+1}(s_n(x)) \in \mathrm{Im}(d'_{n+1} \colon Q_{n+1} \to Q_n),$$

which means it is 0 in $h_n(Q_\bullet)$. Thus, *homotopic maps induce the same map on homology.* We will write $f_\bullet \sim h_\bullet$ to indicate that they are homotopic; letting $g_\bullet = h_\bullet - f_\bullet$ one has $g_\bullet \sim 0$.

Proof of the Comparison Lemma. We have the commutative diagram where the doted arrow, which we denote f_0, is coming from the projectivity of P_0.

$$\begin{array}{ccccc}
 & & P_0 & & \\
 & \overset{f_0}{\swarrow} & \downarrow{\scriptstyle f\epsilon} & & \\
Q_0 & \longrightarrow & B & \longrightarrow & 0
\end{array}$$

(Here and below we write hg for the composition $h \circ g$ to simplify the diagrams.) This constructs the first square in (35) and shows it is commutative. Suppose we have already constructed $f_0, \dots, f_n$ so that the corresponding squares in (35) are commutative. We have then the following information

$$\begin{array}{ccc} P_{n+1} & \xrightarrow{d} & P_n \\ {\scriptstyle f_n d}\downarrow & \swarrow{\scriptstyle f_n} & \\ Q_{n+1} & \longrightarrow & Q_n \end{array}$$

The commutativity of the squares already constructed implies that $\mathrm{Im}(f_n d) \subseteq \mathrm{Ker}(Q_n \to Q_{n-1}) = \mathrm{Im}(Q_{n+1} \to Q_n)$, where we used the exactness of the lower row for the last equality. Thus, we can improve the last diagram, define the dotted arrow (using the projectivity of P_{n+1}) and call it f_{n+1}:

$$\begin{array}{ccccc} & & P_{n+1} & \xrightarrow{d} & P_n \\ & \swarrow{\scriptstyle f_{n+1}} & {\scriptstyle f_n d}\downarrow & \swarrow{\scriptstyle f_n} & \\ Q_{n+1} & \longrightarrow & \mathrm{Im}(Q_{n+1} \to Q_n) & \longrightarrow & 0 \end{array}$$

Focusing on the parallelogram, we see that we have constructed f_{n+1} so that the corresponding square in (35) commutes.

Suppose now that we have two such commuting ladders:

$$\begin{array}{ccccccccccccc} \cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\ & & {\scriptstyle f_2}\downarrow{\scriptstyle h_2} & & {\scriptstyle f_1}\downarrow{\scriptstyle h_1} & & {\scriptstyle f_0}\downarrow{\scriptstyle h_0} & & \downarrow{\scriptstyle f} & & \\ \cdots & \xrightarrow{d'_3} & Q_2 & \xrightarrow{d'_2} & Q_1 & \xrightarrow{d'_1} & Q_0 & \xrightarrow{\epsilon'} & B & \longrightarrow & 0 \end{array}$$

We want to show that $f_\bullet$ is homotopic to $h_\bullet$, after deleting A and B, ϵ and ϵ', from the diagram. Let $g_n = h_n - f_n$. Then we need to construct the following diagram in which

$$g_n = d'_{n+1} s_n + s_{n-1} d_n$$

(where $d_0 = \epsilon$):

$$\begin{array}{ccccccccccc} \cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & 0 \\ & \swarrow{\scriptstyle s_2} & \downarrow{\scriptstyle g_2} & \swarrow{\scriptstyle s_1} & \downarrow{\scriptstyle g_1} & \swarrow{\scriptstyle s_0} & \downarrow{\scriptstyle g_0} & \swarrow{\scriptstyle s_{-1}} & \\ \cdots & \xrightarrow{d'_3} & Q_2 & \xrightarrow{d'_2} & Q_1 & \xrightarrow{d'_1} & Q_0 & \xrightarrow{\epsilon'} & 0 \end{array}$$

We define:

- $s_{-1} = 0$ (no choice);
- s_0 to be a lift of g_0, which exists – we claim – because P_0 is projective. Why is that? $g_0 = h_0 - f_0$ and both h_0 and f_0 lift f in the sense that $\epsilon' h_0 = \epsilon' f_0 = f\epsilon$ and so $\epsilon' g_0 = 0$. This means that the image of g_0 is contained the kernel of ϵ', which is the image of $Q_1 \to Q_0$. So we can look instead at the diagram

$$\begin{array}{ccccc} & & P_0 & & \\ & & {\scriptstyle g_0}\downarrow & & \\ Q_1 & \longrightarrow & \mathrm{Im}(Q_1 \to Q_0) & \longrightarrow & 0 \end{array}$$

  and use the projectivity of P_0. Once defined, it is a formal consequence that

$$g_0 = d'_1 s_0 + s_{-1} d_0;$$

- s_1 to be a lift of $g_1 - s_0 d_1$, which exists – we claim – because P_1 is projective. The argument for that is similar. We have

$$d_1'(g_1 - s_0 d_1) = g_0 d_1 - d_1' s_0 d_1 = g_0 d_1 - (g_0 - s_{-1} d_0) d_1 = s_{-1} d_0 d_1 = 0.$$

We have avoided using that $s_{-1} = 0$ to prove vanishing, relying instead on $d^2 = 0$. This is the way that generalizes for higher indices. The conclusion is that $\mathrm{Im}(g_1 - s_0 d_1) \subseteq \mathrm{Ker}(Q_1 \to Q_0) = \mathrm{Im}(Q_2 \to Q_1)$ and so we have the diagram

$$\begin{array}{ccccc} & & P_1 & & \\ & & \downarrow{\scriptstyle g_1 - s_0 d_1} & & \\ Q_2 & \longrightarrow & \mathrm{Im}(Q_2 \to Q_1) & \longrightarrow & 0 \end{array}$$

and now we can use the projectivity of P_1. Once more, it is a formal consequence that

$$g_1 = d_2' s_1 + s_0 d_1.$$

- Continue to lift inductively $g_2 - s_1 d_2$ to s_2, $g_3 - s_2 d_3$ to s_3, etc.

□

Theorem 35.2.4. *The left derived functor $L_n F$ is an additive functor,*[51] *independent of the choice of resolution used in its definition.*

Proof. Let $f\colon A \to B$ be a morphism of R-modules and choose projective resolutions

$$P_\bullet \to A \to 0, \qquad Q_\bullet \to B \to 0.$$

Using the comparison lemma, we get a morphism of complexes

$$f_\bullet : P_\bullet \to Q_\bullet$$

extending f, and we apply the functor F to get a morphism of complexes

$$Ff_\bullet : FP_\bullet \to FQ_\bullet.$$

This morphism induces homomorphisms on homology

$$\begin{array}{ccc} h_n(FP_\bullet) & \xrightarrow{Ff_n} & h_n(FQ_\bullet) \qquad n \geq 0 \\ \| {\scriptstyle (\mathrm{def})} & & \| {\scriptstyle (\mathrm{def})} \\ (L_n F)(A) & \xrightarrow{(L_n F)(f)} & (L_n F)(B) \end{array}$$

The arrow $(L_n F)(f)$ is *defined* using this diagram. Moreover, this homomorphism $(L_n F)(f)$ is independent on the choice of the morphism $f_\bullet$ extending f. Indeed, if $\tilde{f}_\bullet$ is another extension of f the $f_\bullet \sim \tilde{f}_\bullet$ and so, because F is additive, $Ff_\bullet \sim F\tilde{f}_\bullet$, hence they define the same map on homology.

It is easy to verify that $L_n F$ is an additive functor, meaning $(L_n F)(f+g) = (L_n F)(f) + (L_n F)(g)$, etc. Indeed, if $f_\bullet$ lifts f and $g_\bullet$ lifts g then $f_\bullet + g_\bullet$ (whose n-th component is $f_n + g_n$) lifts $f+g$. At this point, simply running through the definitions yields the identity $(L_n F)(f+g) = (L_n F)(f) + (L_n F)(g)$. Similar arguments work for $(L_n F)(f \circ g) = (L_n F)(f) \circ (L_n F)(g)$ and for showing that $L_n F$ commutes with direct sums.

However, everything we said thus far depends on the chosen resolutions. We want to show that if we take two different projective resolutions of A, say

$$P_\bullet \to A \to 0, \qquad Q_\bullet \to A \to 0,$$

[51] The functor $L_n F$ takes values in S-modules and not just in abelian groups. See Remark 36.0.1.

then we have a *canonical* isomorphism

$$h_n(FP_\bullet) \cong h_n(FQ_\bullet),$$

which is as close as one can get to showing that $(L_nF)(A)$ is independent of the chosen resolution.

Since we have two projective resolutions of A, the Comparison Lemma gives us the following commutative diagram

$$\begin{array}{ccccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle id} & & \\
\cdots & \xrightarrow{d_3'} & Q_2 & \xrightarrow{d_2'} & Q_1 & \xrightarrow{d_1'} & Q_0 & \xrightarrow{\epsilon'} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle g_2} & & \downarrow{\scriptstyle g_1} & & \downarrow{\scriptstyle g_0} & & \downarrow{\scriptstyle id} & & \\
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0
\end{array}$$

Those produce for us, by our formalism, homomorphisms

$$\begin{aligned}
\phi_n &= (L_nF)(id) : h_n(FP_\bullet) \to h_n(FQ_\bullet), \quad \text{defined using } f_\bullet \\
\gamma_n &= (L_nF)(id) : h_n(FQ_\bullet) \to h_n(FP_\bullet), \quad \text{defined using } g_\bullet
\end{aligned}$$

Then, the composition $\gamma_n \circ \phi_n$ is the map $h_n(FP_\bullet) \to h_n(FP_\bullet)$ induced from $g_\bullet \circ f_\bullet$. Note that $g_\bullet \circ f_\bullet$ fits into

$$\begin{array}{ccccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle g_2\circ f_2} & & \downarrow{\scriptstyle g_1\circ f_1} & & \downarrow{\scriptstyle g_0\circ f_0} & & \downarrow{\scriptstyle id} & & \\
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0
\end{array}$$

But we also have the diagram

$$\begin{array}{ccccccccccc}
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle id} & & \downarrow{\scriptstyle id} & & \downarrow{\scriptstyle id} & & \downarrow{\scriptstyle id} & & \\
\cdots & \xrightarrow{d_3} & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \xrightarrow{\epsilon} & A & \longrightarrow & 0
\end{array}$$

Thus, $g_\bullet \circ f_\bullet$ is homotopic to $id_\bullet$ and so induces the identity map $h_n(FP_\bullet) \to h_n(FP_\bullet)$, meaning $\gamma_n \circ \phi_n = id$. Similarly, $\phi_n \circ \gamma_n = id$ and so ϕ_n is an isomorphism. That is, $L_nF(A)$ defined using $P_\bullet \to A \to 0$ is canonically isomorphic to $L_nF(A)$ defined using $Q_\bullet \to A \to 0$. □

Example 35.2.5. If A is projective, $L_nF(A) = 0$ for all $n \geq 1$. Indeed, as we are allowed to choose any projective resolution, we take the resolution $\ldots \to 0 \to 0 \to A \to A \to 0$. We then delete A and apply F to get the complex

$$\ldots \to 0 \to 0 \to 0 \to F(A) \to 0.$$

Its zero homology is $F(A)$ and all its higher homology is 0. Thus, $L_nF(A) = 0$ for all $n \geq 1$.

35.3. **Long exact sequence of the derived functors.** The following theorem is the key theorem. It is the reason for everything we have done so far.

Theorem 35.3.1 (Long exact sequence of the derived functors)**.** *Let F be a right-exact covariant functor, $F\colon R\text{−}\mathbf{Mod} \to S\text{−}\mathbf{Mod}$, and let*

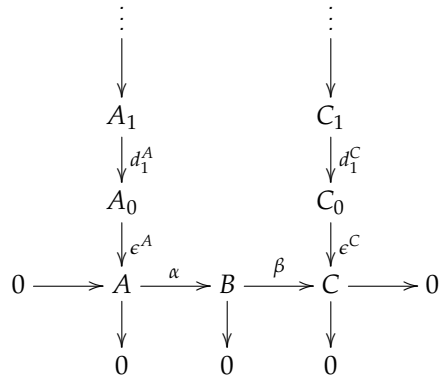
$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$$

be an exact sequence of R-modules. There is a long exact sequence of S-modules

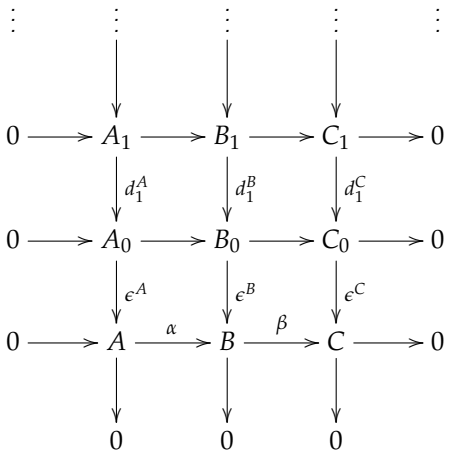
$$\ldots \longrightarrow (L_2F)(C) \longrightarrow (L_1F)(A) \longrightarrow (L_1F)(B) \longrightarrow (L_1F)(C) \longrightarrow F(A) \longrightarrow F(B) \longrightarrow F(C) \longrightarrow 0.$$

Proof. We will need the following lemma.

Lemma 35.3.2 (Horseshoe Lemma)**.** *Given a partial projective resolution of an exact row*



there is a projective resolution of B so that we get exact rows.



Proof. (Horseshoe Lemma) The construction is inductive. We first define

$$B_0 = A_0 \oplus C_0$$

with the obvious inclusion $A_0 \hookrightarrow B_0$ and surjection $B_0 \twoheadrightarrow C_0$. Note that B_0 is projective. Since C_0 is projective, there is a morphism

$$f\colon C_0 \to B, \quad \beta \circ f = \epsilon^C.$$

Let

$$\epsilon^B\colon B_0 \to B, \quad \text{such that } \epsilon^B(a_0, c) = \alpha\epsilon^A(a_0) + f(c).$$

It is straightforward to verify that we get now a commutative diagram with exact rows (the bottom two rows are exact and one can conclude the exactness of the top row from the Snake Lemma):

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathrm{Ker}(\epsilon^A) & \longrightarrow & \mathrm{Ker}(\epsilon^B) & \longrightarrow & \mathrm{Ker}(\epsilon^C) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & A_0 & \longrightarrow & B_0 & \longrightarrow & C_0 & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \epsilon^A} & & \downarrow{\scriptstyle \epsilon^B} & & \downarrow{\scriptstyle \epsilon^C} & & \\
0 & \longrightarrow & A & \xrightarrow{\alpha} & B & \xrightarrow{\beta} & C & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

From this diagram we conclude the following diagram

$$\begin{array}{ccccccccc}
 & & A_1 & & & & C_1 & & \\
 & & \downarrow{\scriptstyle d_1^A} & & & & \downarrow{\scriptstyle d_1^C} & & \\
0 & \longrightarrow & \mathrm{Ker}(\epsilon^A) & \longrightarrow & \mathrm{Ker}(\epsilon^B) & \longrightarrow & \mathrm{Ker}(\epsilon^C) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

We can repeat the first step arguments and get a diagram with exact rows such that B_1 is projective,

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A_1 & \longrightarrow & B_1 & \longrightarrow & C_1 & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle d_1^A} & & \downarrow{\scriptstyle d_1^B} & & \downarrow{\scriptstyle d_1^C} & & \\
0 & \longrightarrow & \mathrm{Ker}(\epsilon^A) & \longrightarrow & \mathrm{Ker}(\epsilon^B) & \longrightarrow & \mathrm{Ker}(\epsilon^C) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

Now repeat the second step arguments, and so on. □

Using the lemma we now prove the theorem. Given an exact sequence $0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0$, we first find projective resolutions $P_\bullet \to A \to 0$ and $P''_\bullet \to C \to 0$. Using the Horseshoe Lemma we find a projective resolution $P'_\bullet \to B \to 0$ so that we get an exact sequence of complexes

$$0 \to P_\bullet \to P'_\bullet \to P''_\bullet \to 0.$$

That means that for every n,

$$0 \to P_n \to P'_n \to P''_n \to 0$$

is exact. Applying F we get a complex of sequences

$$0 \to FP_\bullet \to FP'_\bullet \to FP''_\bullet \to 0.$$

Meaning, that for every n, we have a complex

$$0 \to FP_n \to FP'_n \to FP''_n \to 0.$$

However, since F is right-exact, in fact this is almost a short exact sequence; the only point missing is that we don't know yet that $FP_n \to FP'_n$ is injective. We use projectivity to prove that.

Firstly, since P''_n is projective, the diagram

$$\begin{array}{ccccc} & & P''_n & & \\ & \swarrow & \downarrow{\scriptstyle id} & & \\ P'_n & \longrightarrow & P''_n & \longrightarrow & 0 \end{array}$$

shows that the sequence $0 \to P_n \to P'_n \to P''_n \to 0$ splits and $P'_n \cong P_n \oplus P''_n$ (cf. proof of Theorem 31.1.3). Since F is additive, we find that

$$FP'_n \cong FP_n \oplus FP''_n$$

and thus $0 \to FP_n \to FP'_n \to FP''_n \to 0$ is exact. We therefore have a SES of complexes

$$0 \to P_\bullet \to P'_\bullet \to P''_\bullet \to 0.$$

Applying Theorem 34.3.2 we find a long exact sequence in homology, as claimed.

□

36. Tor

Given two R-modules, $A \in \mathbf{Mod_R}$ and $B \in {}_\mathbf{R}\mathbf{Mod}$, there are two things we can do. We can think about the functor $A \otimes_R (-)$ from ${}_\mathbf{R}\mathbf{Mod}$ to abelian groups (and even to R-modules if R is commutative). As such it has left derived functors $L_nA \otimes_R (-)$ (that is, L_nF, where $F(-) = A \otimes_R (-)$) that we can evaluate on any module in ${}_\mathbf{R}\mathbf{Mod}$, and in particular on B. We get

$$(L_nA \otimes_R (-))(B) = h_n(A \otimes P^B_\bullet),$$

where $P^B_\bullet \to B \to 0$ is a projective resolution of B.

On the other hand, we can think about the functor $(-) \otimes_R B$ from $\mathbf{Mod}_R$ to abelian groups (and even to R-modules if R is commutative). As such it has left derived functors $L_n(-) \otimes_R B$ (that is, L_nG, where $G(-) = (-) \otimes_R B$) that we can evaluate on any module in $\mathbf{Mod}_R$, and in particular on A. We get

$$(L_n(-) \otimes_R B)(A) = h_n(P^A_\bullet \otimes B),$$

where $P^A_\bullet \to A \to 0$ is a projective resolution of A.

It is a theorem that we get the same result; there is a natural isomorphism

$$h_n(A \otimes P^B_\bullet) \cong h_n(P^A_\bullet \otimes B).$$

We will not prove this theorem. It can be found in Rotman's book. One denotes this common value as

$$\mathrm{Tor}_n(A,B) = (L_nA \otimes_R (-))(B) = (L_n(-) \otimes_R B)(A)$$

Note that it follows that $\mathrm{Tor}_n(A,B)$ is a **bifunctor**: for A fixed it is a covariant additive functor in B, and for B fixed it is a covariant additive functor in A.

The main results we have so far concerning this construction are:

(1) $\mathrm{Tor}_0(A,B) = A \otimes B$.
(2) If A, or B, are projective then $\mathrm{Tor}_n(A,B) = \{0\}$ for $n \geq 1$.
(3) Given a short exact sequence $0 \to B' \to B \to B'' \to 0$ we have a long exact sequence

$$\ldots \longrightarrow \mathrm{Tor}_1(A,B'') \longrightarrow A \otimes B' \longrightarrow A \otimes B \longrightarrow A \otimes B'' \longrightarrow 0.$$

And, given a short exact sequence $0 \to A' \to A \to A'' \to 0$ we have a long exact sequence

$$\ldots \longrightarrow \mathrm{Tor}_1(A'',B) \longrightarrow A' \otimes B \longrightarrow A \otimes B \longrightarrow A'' \otimes B \longrightarrow 0.$$

Remark 36.0.1. Suppose that $M \in {}_{\mathbf{R}}\mathbf{Mod}_{\mathbf{S}}$ is a bimodule. We can think about every element $s \in S$ as defining an R-module endomorphism $M \to M, m \mapsto ms$. Let $A \in \mathbf{Mod}_{\mathbf{R}}$. Since the left derived functors of $A \otimes_R (-)$ are functors, this homomorphism induces an endomorphism of $\mathrm{Tor}_n(A, M)$ and we conclude that $\mathrm{Tor}_n(A, M)$ takes value in $\mathbf{Mod}_{\mathbf{S}}$ if $M \in {}_{\mathbf{R}}\mathbf{Mod}_{\mathbf{S}}$. In a similar way, if $A \in {}_{\mathbf{S}}\mathbf{Mod}_{\mathbf{R}}$, $\mathrm{Tor}_n(A, M) \in {}_{\mathbf{S}}\mathbf{Mod}$. (The fact that the derived functors are additive guarantees in both cases that the module axioms hold for $\mathrm{Tor}_n(A, M)$.

36.1. **Tor and flatness.** We can use the functor Tor_1 to characterize flat modules. We state the result of right R-modules but, with the obvious adjustments, it holds of left R-modules too.

Theorem 36.1.1. *If A is a flat R-module, $\mathrm{Tor}_n(A, B) = \{0\}$ for all $n \geq 1$ and all left R-modules B. Conversely, if $\mathrm{Tor}_1(A, B) = 0$ for all B, then A is flat.*

Proof. Suppose that A is flat and $P_\bullet^B \to B \to 0$ is a projective resolution of B. Then $A \otimes P_\bullet^B$ is still exact in all positive degrees (since we have deleted $A \otimes B$ there is no reason for the sequence $A \otimes P_1^B \to A \otimes P_0^B \to 0$ to be exact). Therefore,

$$(L_n A \otimes (-))(B) = h_n(A \otimes P_\bullet^B) = 0, \qquad \forall n \geq 1.$$

Conversely, given that $\mathrm{Tor}_1(A, B) = 0$ for all B, we need to show that $A \otimes (-)$ is an exact functor. But, if $0 \to B' \to B \to B'' \to 0$ is exact, so is

$$0 = \mathrm{Tor}_1(A, B'') \to A \otimes B' \to A \otimes B \to A \otimes B'' \to 0.$$

□

36.2. **Tor and torsion.** The choice of notation Tor comes from the word *torsion* and in this section we prove a theorem that explains this connection.

Let R be a commutative integral domain, $Q = \mathrm{Frac}(R)$ the fraction field of R and $K = Q/R$. Good examples to keep in mind are $R = \mathbb{Z}, Q = \mathbb{Q}$, or $R = \mathbb{C}[x], Q = \mathbb{C}(x)$, but these examples are special by virtue of being PID's.

For an R-module A denote by A_t the **torsion** submodule of A:

$$A_t = \{a \in A : ra = 0, \text{for some } r \in R, r \neq 0\}.$$

It is easy to see that $A \mapsto A_t$ is a covariant additive functor on R-modules. We say that A is **torsion** if $A = A_t$. We say that A is **torsion-free** if $A_t = 0$.

Note that Q is a flat R-module because the functor $(-) \otimes_R Q$ is the same as localizing at the multiplicative set $S = R - \{0\}$, which we proved is an exact functor. Thus,

(1) Since R is free it is projective and so $\mathrm{Tor}_n(R, A) = 0$ for all $n \geq 1$.
(2) Since Q is flat, $\mathrm{Tor}_n(Q, A) = 0$ for all $n \geq 1$ by Theorem 36.1.1.

By taking the long exact sequence associated to (36) for the operation of $\otimes_R A$ we get

$$\begin{array}{ccccc} \mathrm{Tor}_n(Q, A) & \longrightarrow & \mathrm{Tor}_n(K, A) & \longrightarrow & \mathrm{Tor}_{n-1}(R, A), \quad n \geq 2. \\ \| & & & & \| \\ 0 & & & & 0 \end{array}$$

Thus,

(3) For any R-module A, $\mathrm{Tor}_n(K, A) = 0$, for $n \geq 2$.

Theorem 36.2.1. *For R an integral domain, the functors $\mathrm{Tor}_1(K,-)$ and $(-)_t$ are naturally equivalent.*

Proof. The exact sequence

$$0 \longrightarrow R \longrightarrow Q \longrightarrow K \longrightarrow 0\,, \tag{36}$$

gives an exact sequence

$$\begin{array}{ccccccc} \mathrm{Tor}_1(Q,A) & \longrightarrow & \mathrm{Tor}_1(K,A) & \longrightarrow & R\otimes_R A & \longrightarrow & Q\otimes_R A \\ \| & & \| & & \| & & \| \\ 0 & \longrightarrow & \mathrm{Tor}_1(K,A) & \longrightarrow & A & \longrightarrow & A[S^{-1}] \end{array}$$

This implies that $\mathrm{Tor}_1(K,A) = \mathrm{Ker}(A \to A[S^{-1}]) = A_t$ (see Page 23 and Lemma 3.2.2). □

36.3. **Tor calculations for a PID.** Let R be a PID and M, N finitely generated R-modules. As, in general,

$$(L_nF)(A \oplus B) \cong (L_nF)(A) \oplus (L_nF)(B),$$

the structure theorems for finitely generated modules over a PID, reduces the computation of $\mathrm{Tor}_n(M,N)$ to the case where $M = R/I, N = R/J$ for some ideals I, J of R. If I or J are the zero ideals, M or N are R and the situation is easy because R is a flat R-module and so all $\mathrm{Tor}_n(M,N)$ vanish for $n \geq 1$. In general, we have the following result.

Theorem 36.3.1. *Let R be a PID and I, J ideals of R.*

(1) $\mathrm{Tor}_0(R/I, R/J) \cong R/(I+J)$.
(2) $\mathrm{Tor}_1(R/I, R/J) \cong (I \cap J)/IJ$.
(3) $\mathrm{Tor}_n(R/I, R/J) = 0$, $n \geq 2$.

Proof. If either of the ideals is 0, the statements hold true as explained above. Thus, the interesting case is when neither of the ideals is the zero ideal and this is what we compute next.

Assume that I and J are non-zero ideals. Let $J = \langle j \rangle$ and let $[j]$ denote the multiplication-by-j map. Then, the sequence

$$0 \longrightarrow R \xrightarrow{[j]} R \longrightarrow R/J \longrightarrow 0$$

is exact and, in fact, forms a projective resolution of R/J. Tensor it with R/I and delete the first term, as usual. By definition, $\mathrm{Tor}_n(R/I, R/J)$ is the n-th homology of

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow R/I \xrightarrow{[j]} R/I \longrightarrow 0$$

So:

- $\mathrm{Tor}_0(R/I, R/J) = R/I \otimes R/J \cong (R/I)/([j](R/I)) = (R/I)/((I+J)/I) \cong R/(I+J)$.
- $\mathrm{Tor}_1(R/I, R/J) = \mathrm{Ker}([j] : R/I \to R/I) \cong \frac{1}{j}(I \cap J)/I \cong I \cap J/IJ$.
- $\mathrm{Tor}_n(R/I, R/J) = 0$ for $n \geq 0$.

□

Remark 36.3.2. One can prove a more general result. For a general ring R, a right ideal I and a left ideal J, the following holds:

(1) $\mathrm{Tor}_1(R/I, R/J) \cong (I \cap J)/IJ$.
(2) $\mathrm{Tor}_2(R/I, R/J) \cong \mathrm{Ker}(I \otimes J \to IJ)$.
(3) $\mathrm{Tor}_{n+2}(R/I, R/J) \cong \mathrm{Tor}_n(I, J), n > 0$.

36.4. **Serre's intersection multiplicity.** The discussion in this section assumes familiarity with algebraic geometry. But, it will be valuable if you simply think about the indicated special cases.

Let X be a regular scheme,[52] and let Y, Z be two closed subschemes of X defined by ideal sheaves I, J and intersecting properly[53]. The intersection multiplicity $m(x, Y, Z)$ of Y and Z at a point $x \in Y \cap Z$ is defined as

$$\text{Serre's formula:} \quad m(x,Y,Z) = \sum_{i\geq 0}(-1)^i \cdot \text{length}_{\mathcal{O}_{X,x}} \text{Tor}_i^{\mathcal{O}_{X,x}}(\mathcal{O}_{X,x}/I_x, \mathcal{O}_{X,x}/J_x)$$

Here $\mathcal{O}_{X,x}$ is the local ring of X at x, I_x, J_x the stalks at x, and the Tor's and lengths are calculated in the category of $\mathcal{O}_{X,x}$-modules.[54]

As the name suggests, this quantity should be a measure of the "order of contact" between Y and Z at the point x. Note that the first term (cf. p. 17) is

$$\text{length}_{\mathcal{O}_{X,x}}(\mathcal{O}_{X,x}/(I_x + J_x)),$$

which is what one comfortable enough with algebraic geometry would expect. The additional terms serve as correction terms, which are necessary once one examines slightly complicated examples.

Consider the case of two curves Y, Z on a regular surface X that do not have a common component. The curves are defined by the ideals I and J. Locally at every point x on the surface we have

$$I_x = (f), \quad J_x = (g),$$

where we use I_x, J_x to denote the localizations of the ideals at the prime ideal corresponding to x. Note that it is not possible for f or g to be equal to 0. The same technique used to calculate Tor for a PID can be employed here. We have an exact sequence

$$0 \longrightarrow \mathcal{O}_x \xrightarrow{[f]} \mathcal{O}_x \longrightarrow \mathcal{O}_x/I_x \longrightarrow 0$$

which is a projective resolution of $\mathcal{O}_x/I_x$ in the category of $\mathcal{O}_x$-modules. We see that

$$\text{Tor}_n^{\mathcal{O}_x}(\mathcal{O}_x/I_x, -) = 0, \forall n \geq 2,$$

while

$$\text{Tor}_1^{\mathcal{O}_x}(\mathcal{O}_x/I_x, \mathcal{O}_x/J_x) \cong (I_x \cap J_x)/I_xJ_x.$$

Although the local ring at x, $\mathcal{O}_x$, is typically not be a PID, it is a regular local ring and hence a UFD, by a classical theorem in commutative algebra. Thus, it is meaningful to say that f and g have no common factor and one finds that $I_x \cap J_x = (fg) = I_xJ_x$. Consequently, $\text{Tor}_1^{\mathcal{O}_x}(\mathcal{O}_x/I_x, \mathcal{O}_x/J_x) = 0$ and Serre's multiplicity formula is simply

$$m(x,Y,Z) = \text{length}_{\mathcal{O}_{X,x}}(\mathcal{O}_{X,x}/(I_x + J_x)).$$

No higher Tor terms are needed for curves on a regular surface.

---

[52] Such as $\text{Spec}(\mathbb{C}[x,y])$, the complex affine plane.

[53] Say $Y = \text{Spec}(\mathbb{C}[x,y]/I), Z = \text{Spec}(\mathbb{C}[x,y]/J)$, but consider even the simpler setting where $I = (f(x,y)), J = (g(x,y))$ where $f(x,y), g(x,y)$ are irreducible polynomials and $I \neq J$. Then Y and Z are the curves in the affine plane defined by $f(x,y) = 0$ and $g(x,y) = 0$, respectively.

[54] Say $x = 0$ then this ring $R = \mathcal{O}_{X,x}$ is the localization of $\mathbb{C}[x,y]$ at the maximal ideal (x,y) and it is a UFD. If $I = (f(x,y)), J = (g(x,y))$ then $I_x = Rf, J_x = Rg$. R is a local ring but not a PID, so the calculation of the Tor_n is a bit more complicated, but see the remark above. Using that R is a UFD, one can show in this case that if f and g are irreducible and do not differ by a scalar, all the higher Tor's vanish and we find

$$m(x,Y,Z) = \text{length}_R(R/(f,g)).$$

This vanishes if either f or g do not pass through 0.

To illustrate Serre's formula, take the curves $Y = (y), Z = (x)$ in the plane $\mathbb{A}^2$. Their intersection multiplicity at the origin $P = (0,0)$, corresponding to the ideal (x,y), is

$$m(P,Y,Z) = \text{length}_{\mathcal{O}_P}(\mathcal{O}_P/(I_P + J_P)) = \text{length}(k[x,y]_{\langle x,y\rangle}/\langle x,y\rangle) = \text{length}(k) = 1.$$

On the other hand, if we take the curves $Y = (x), Z = (y^2 - x^3)$, we find that

$$m(P,Y,Z) = \text{length}(k[x,y]_{\langle x,y\rangle}/\langle x, y^2 - x^3\rangle) = \text{length}(k[x,y]_{\langle x,y\rangle}/\langle x, y^2\rangle) = 2.$$

(For the last equality we used the filtration $\mathcal{O}_P \supset \langle x,y\rangle \supset \langle x,y^2\rangle$ that has simple quotients.) Thus, the curves have order of contact 2 at the origin.

To see why the $\mathbf{Tor}$ terms are necessary, one needs to go to slightly more complicated examples. Let Y be the union of the planes $\langle x,y\rangle$ and $\langle z,w\rangle$ in $\mathbb{A}^4$. Thus, Y is defined by the ideal

$$\langle x,y\rangle \cap \langle z,w\rangle = \langle xz, yz, xw, yw\rangle.$$

Let Z be the diagonal plane defined by $\langle x - z, y - w\rangle$. As Z meets each of the planes forming Y at one point $P = (0,0,0,0)$ (intuitively, or using the multiplicity formula) we should expect

$$m(P,Y,Z) = 1 + 1 = 2.$$

However, if we let A be the local ring of $\mathbb{A}^4$ at P and we calculate the length of

$$A/(\langle xz, yz, xw, yw\rangle + \langle x - z, y - w\rangle) = A/\langle xz, yz, xw, yw, x - z, y - w\rangle \cong k[x,y]/(x^2, xy, y^2),$$

we find that its length is 3. The higher $\mathbf{Tor}$ terms are then providing the needed correction. Only $\mathbf{Tor}_1$ is not trivial, being of length 1; the rest are 0. We do not compute them here – the computation is rather subtle – but we do remark that Remark 36.3.2 is very useful in the calculations.

Remark 36.4.1. In spite of the elegance of Serre's formula, there were (and still are) formidable difficulties associated with it, especially when one relaxes the hypotheses. To make sense of the formula one would like to know that if R is a regular local ring (such as $\mathcal{O}_{X,x}$) and I, J are prime ideals of it then

$$\dim(R/I) + \dim(R/J) \leq \dim R.$$

This was proved by Serre. One would also like to know that:

(1) $m(x,Y,Z) \geq 0$, which was proved by Serre in the case we are considering, and by Ofer Gabber in general.
(2) $m(x,Y,Z) = 0$ if $\dim(R/I) + \dim(R/J) < \dim R$. This was proven by Paul Roberts and, independently, by Gillet and Soulé.
(3) $m(x,Y,Z) > 0$ if $\dim(R/I) + \dim(R/J) = \dim R$. This remains open in general, although in many cases of interest it is known.

37. Derived functors – II

In this section we will define the **right derived functors**

$$R^n F, \quad n \geq 0,$$

of a left-exact covariant functor F. The theory works for a functor $F\colon \mathbb{A} \to \mathbb{B}$, where $\mathbb{A}, \mathbb{B}$ are abelian categories and $\mathbb{A}$ has enough injectives. But to make things more concrete we will assume that both categories are categories of modules over a ring.

37.1. **Three key examples.** Here are three examples everyone should care about.

(1) F is the functor $\mathrm{Hom}_R(M,-)$ from the category of R-modules to abelian groups (resp., to R-modules if R is commutative). We have seen that this is a left-exact covariant functor. Recall that the module M is called projective if this functor is exact and, in general, this isn't the case. To define its right-derived functors, to be denoted $\mathrm{Ext}^n(M,-)$, we will use that the category of R-modules has enough injectives.

(2) Let $(X,\mathcal{O}_X)$ be a locally ringed space, such as the spectrum of a commutative ring R, or a complex manifold. The category of sheaves $\mathscr{F}$ on X has enough injectives. The global sections functor, usually denoted by
$$\Gamma(X,\mathscr{F}) := \mathscr{F}(X),$$
is a left-exact covariant functor. Meaning, if
$$0 \to \mathscr{A} \to \mathscr{B} \to \mathscr{C} \to 0,$$
is an exact sequence of sheaves, then the sequence of abelian groups
$$0 \to \Gamma(X,\mathscr{A}) \to \Gamma(X,\mathscr{B}) \to \Gamma(X,\mathscr{C})$$
is exact but the last arrow is not always surjective. For example, if X is a complex manifold, we have the exact sequence
$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathcal{O}_X \xrightarrow{e^{2\pi i(-)}} \mathcal{O}_X^\times \longrightarrow 0,$$
where $\mathbb{Z}$ is the sheaf of locally-constant functions with integer values, $\mathcal{O}_X$ is the sheaf of holomorphic functions and $\mathcal{O}_X^\times$ is the sheaf of invertible (equivalently, non-vanishing) holomorphic functions. The fact that
$$\mathcal{O}_X(X) \to \mathcal{O}_X^\times(X)$$
is usually not surjective has to do with the fact that usually there is no well-defined logarithm of a function. This problem already arises for $X = \mathbb{C} - \{0\}$ and the function $f(x) = x$ on $\mathbb{C} - \{0\}$. There isn't a well-defined function $\log(x)$ on the whole of $\mathbb{C} - \{0\}$; one could say that this is the origin of the theory of Riemann surfaces.

The category of sheaves on X has enough injective (but apparently doesn't have enough projectives). This is a lucky break; without it, geometry would still be in the stone-age.

(3) Let G be a finite group, typically non-abelian, and let A,B,C be $\mathbb{Z}[G]$-modules. That is, A,B,C are abelian groups equipped with an action of G that respects addition. For example, A,B,C could be complex representations of G. Or, A,B,C could be finite abelian groups and G a group that maps to $\mathrm{Aut}(A),\mathrm{Aut}(B),\mathrm{Aut}(C)$, and so-on.

Assume that we have an exact sequence of $\mathbb{Z}[G]$-modules,
$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0.$$
We then pass to invariants: $A^G := \{a \in A : ga = a, \forall g \in G\}$ and we get an exact sequence
$$0 \longrightarrow A^G \longrightarrow B^G \longrightarrow C^G,$$
where the last arrow need not be surjective. A very simple example is provided by $G = \{\pm 1\}$ acting on any abelian group (written additively) by $a \mapsto -a$. Then, the sequence of invariants is
$$0 \longrightarrow A[2] \longrightarrow B[2] \longrightarrow C[2],$$
where $A[2]$ means the elements a of A such that $2a = 0$. Take the case
$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{x\mapsto 2x} \mathbb{Z}/4\mathbb{Z} \xrightarrow{x\mapsto x \bmod 2} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

Taking invariants we find

$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \xrightarrow{x\mapsto 2x} 2\mathbb{Z}/4\mathbb{Z} \xrightarrow{x\mapsto x \bmod 2} \mathbb{Z}/2\mathbb{Z}$$

where the last arrow is the zero map, and so is not surjective.

37.2. **The right derived functors of a covariant left-exact functor.** Roughly speaking, the theory is constructed the same way we have constructed it for right-exact functors, only that arrows get reversed and one uses injective resolutions now. Here is a more detailed sketch:

Let $F\colon R\text{–}\mathbf{Mod} \to S\text{–}\mathbf{Mod}$ be a left-exact covariant additive functor.

- Given an R-module A, choose an *injective* resolution $I_\bullet$ of A,

$$0 \longrightarrow A \xrightarrow{\epsilon} I_0 \xrightarrow{d_0} I_1 \xrightarrow{d_1} I_2 \xrightarrow{d_2} \cdots$$

  Meaning, this is an exact sequence of R-modules and each I_j is an injective R-module. [55]

- Apply F to $I_\bullet$ to obtain a complex

$$0 \longrightarrow FI_0 \xrightarrow{Fd_0} FI_1 \xrightarrow{Fd_1} FI_2 \xrightarrow{Fd_2} \cdots$$

- Take the homology to get the **right derived functors** of F,

$$(R^nF)(A) := h_n(FI_\bullet) = \frac{\mathrm{Ker}(FI_n \to FI_{n+1})}{\mathrm{Im}(FI_{n-1} \to FI_n)}, \quad n \geq 0$$

Then the following holds:

(1) $R^0F(A) \cong FA$ (because this is $\mathbf{Ker}(FI_0 \to FI_1)$, which, by virtue F being left-exact, is the image of $FA \xrightarrow{F\epsilon} FI_0$, which is isomorphic to FA).

(2) There is a Comparison Lemma for injective resolutions that proves, using a argument similar to the one we gave for right-exact functors, that:
   (a) $(R^nF)(A)$ is independent of the resolution up to canonical isomorphisms;
   (b) A homomorphism $f\colon A \to B$ induces a homomorphism $(R^nF)(f)\colon (R^nF)(A) \to (R^nF)(B)$, thus making each R^nF into a covariant additive functor $R\text{–}\mathbf{Mod} \to S\text{–}\mathbf{Mod}$.

(3) Given a long exact sequence of R-modules $0 \to A \to B \to C \to 0$ there is a long exact sequence of S-modules:

$$0 \longrightarrow F(A) \longrightarrow F(B) \longrightarrow F(C) \longrightarrow R^1F(A) \longrightarrow R^1F(B) \longrightarrow R^1F(C) \longrightarrow R^2F(A) \longrightarrow \cdots$$

The last statement is proven by first proving a Horseshoe Lemma for injective resolutions and then using the long exact sequence in homology (that we already proved in the generality needed). All that is not hard; it is a rather straightforward adaptation of what we did.

For the case $R^n\mathrm{Hom}_R(M,-)$ we introduce special notation:

$$\mathrm{Ext}^n(M,-) := R^n\mathrm{Hom}_R(M,-)$$

[55] It seems to me more transparent to index the objects with increasing indices than to insist on decreasing indices as in our original definition of a complex.

and its worth remembering how it's defined: Given A choose an injective resolution $0 \to A \to I_\bullet$ and let

$$\mathrm{Ext}^n(M,-) = h_n(\mathrm{Hom}(M, I_\bullet)).$$

There is also special notation for the functors $\Gamma(X,-)$ and $(-)^G$:

$$H^n(X,-) = R^n\Gamma(X,-), \qquad H^n(G,-) = R^n(-)^G.$$

The first is referred to as **sheaf cohomology** and the second as **group cohomology**.

37.3. **Yet another variant (that cannot be ignored).** We can consider a *contravariant* additive left-exact functor F and we can define its right derived functor. It's the usual story: Given a module A choose a projective resolution for A,

$$\cdots \longrightarrow P_2^A \longrightarrow P_1^A \longrightarrow P_0^A \longrightarrow A \longrightarrow 0$$

Remove A and apply F (it's contravariant!) to get a complex

$$0 \longrightarrow FP_0^A \longrightarrow FP_1^A \longrightarrow FP_2^A \longrightarrow \cdots$$

and take its homology to get the **right derived functors** of F,

$$R^nF(A), \quad n \geq 0.$$

In particular, this applies to the functor $\mathrm{Hom}_R(-,N)$.

As in the case of Tor, one can play the following game. Let A and B be R-modules then we can look at the two abelian groups

$$(R^n\mathrm{Hom}_R(A,-))(B) = h_n(\mathrm{Hom}_R(A, I_\bullet^B)), \qquad (R^n\mathrm{Hom}_R(-,B))(A) = h_n(\mathrm{Hom}(P_\bullet^A, B)).$$

It is a non-trivial theorem that those two abelian groups are canonically isomorphic (see Rotman, Theorem 7.8); one denotes their common value $\mathrm{Ext}^n(A,B)$. Thus,

$$\mathrm{Ext}^n(A,B) \cong h_n(\mathrm{Hom}(P_\bullet^A, B)) \cong h_n(\mathrm{Hom}_R(A, I_\bullet^B)) \quad .$$

The following properties are easy to prove from the definitions (and same ideas used for Tor):

(1) $\mathrm{Ext}^n(A,B)$ is a bifunctor – covariant in B, contravariant in A.
(2) $\mathrm{Ext}^0(A,B) = \mathrm{Hom}_R(A,B)$.
(3) $\mathrm{Ext}^n(A,B) = 0$ for $n \geq 1$ and any A, if B is *injective*.
(4) $\mathrm{Ext}^n(A,B) = 0$ for $n \geq 1$ and any B, if A is *projective*.

Remark 37.3.1. Here is a summary of which resolutions and which derived functors are defined in each case.

Type of functor	1/2 exactness	Resolution used	Derived functors	Main examples
Covariant	right exact	projective	left derived L_nF	$\otimes_R M$
	left exact	injective	right derived R_nF	$\mathrm{Hom}_R(M,-)$, $\Gamma(X,-)$, A^G
Contravariant	right exact	injective	left derived L_nF	do not arise often
	left exact	projective	right derived R_nF	$\mathrm{Hom}_R(,-,M)$

37.4. **Examples.** The calculation of $\mathrm{Ext}^n(A,B)$, where A and B are finitely generated abelian groups, reduces to the case of cyclic groups and $\mathrm{Ext}^n(A,B)$ is an additive functor in both A and B. The same discussion, with obvious modifications, would be valid for finitely generated modules over a PID.

Lemma 37.4.1. *Let a,b be positive integers. Then*

(1) $\mathrm{Ext}^n(\mathbb{Z},\mathbb{Z}) = \begin{cases} \mathbb{Z} & n=0, \\ \{0\} & n \geq 1. \end{cases}$

(2) $\mathrm{Ext}^n(\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \begin{cases} \mathbb{Z}/b\mathbb{Z} & n=0, \\ \{0\} & n \geq 1. \end{cases}$

(3) $\mathrm{Ext}^n(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = \begin{cases} \{0\} & n=0 \\ \mathbb{Z}/a\mathbb{Z} & n=1, \\ \{0\} & n>1. \end{cases}$

(4) $\mathrm{Ext}^n(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \begin{cases} \mathbb{Z}/(a,b)\mathbb{Z} & n=0 \\ \mathbb{Z}/(a,b)\mathbb{Z} & n=1, \\ \{0\} & n>1. \end{cases}$

Proof. As $\mathbb{Z}$ is projective, we get the vanishing of $\mathrm{Ext}^n(\mathbb{Z},B)$ for $n \geq 1$ for any abelian group B and, in particular, $\mathrm{Ext}^n(\mathbb{Z},\mathbb{Z}) = 0$ for $n \geq 1$. $\mathrm{Ext}^0(\mathbb{Z},\mathbb{Z}) = \mathrm{Ext}^0(\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) = \mathbb{Z}$, and $\mathrm{Ext}^0(\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \mathrm{Hom}(\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \mathbb{Z}/b\mathbb{Z}$. Thus, we have proven (1) and (2).

Consider the statements in (3). First, $\mathrm{Ext}^0(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = \mathrm{Hom}(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = 0$. To calculate the other Ext we consider the short exact sequence for multiplication by a,

$$0 \longrightarrow \mathbb{Z} \xrightarrow{[a]} \mathbb{Z} \longrightarrow \mathbb{Z}/a\mathbb{Z} \longrightarrow 0, \tag{37}$$

and the functor $\mathrm{Hom}(-,\mathbb{Z})$. There are two ways to proceed: we can view the sequence as a projective resolution $P_\bullet$ of $\mathbb{Z}/a\mathbb{Z}$ and by definition then $\mathrm{Ext}^n(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = h_n(\mathrm{Hom}(P_\bullet,\mathbb{Z}))$. That is, the homology of

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) & \longrightarrow & \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) & \longrightarrow & \mathrm{Hom}(0,\mathbb{Z}) & \longrightarrow & \dots \\ & & \| & & \| & & \| & & \\ 0 & \longrightarrow & \mathbb{Z} & \longrightarrow & \mathbb{Z} & \longrightarrow & 0 & \longrightarrow & \end{array}$$

Under the identifications the map $\mathbb{Z} \to \mathbb{Z}$ is multiplication by a. We thus find

$$\mathrm{Ext}^0(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = 0, \quad \mathrm{Ext}^1(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = \mathbb{Z}/a\mathbb{Z}, \quad \mathrm{Ext}^n(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) = 0, \forall n \geq 2.$$

Another option, is to take the long exact sequence in homology associated to (37) and the functor $\mathrm{Hom}(-,\mathbb{Z})$, and get

$$\begin{array}{ccccccccccccc} 0 & \to & \mathrm{Hom}(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) & \to & \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) & \to & \mathrm{Hom}(\mathbb{Z},\mathbb{Z}) & \to & \mathrm{Ext}^1(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}) & \to & \mathrm{Ext}^1(\mathbb{Z},\mathbb{Z}) & \to & \dots \\ & & \| & & \| & & \| & & \| & & \| & & \\ 0 & \longrightarrow & 0 & \longrightarrow & \mathbb{Z} & \xrightarrow{[a]} & \mathbb{Z} & \longrightarrow & ? & \longrightarrow & 0 & \longrightarrow & \dots \end{array}$$

and use the computations of $\mathrm{Ext}^n(\mathbb{Z},\mathbb{Z})$ to find the value of the question marks.

Finally, consider the last statement. Once more we use the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{[a]} \mathbb{Z} \longrightarrow \mathbb{Z}/a\mathbb{Z} \longrightarrow 0, \tag{38}$$

which we view as projective resolution of $\mathbb{Z}/a\mathbb{Z}$. Applying the functor $\mathrm{Hom}(-,\mathbb{Z}/b\mathbb{Z})$ and omitting the first term we get the complex

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Hom}(\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) & \xrightarrow{[a]} & \mathrm{Hom}(\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) & \longrightarrow & 0 & \longrightarrow & 0 \longrightarrow \dots \\ & & \| & & \| & & & & \\ & & \mathbb{Z}/b\mathbb{Z} & \xrightarrow{[a]} & \mathbb{Z}/b\mathbb{Z} & & & & \end{array}$$

Let $d = \gcd(a,b)$. Then, from the sequence, $\mathrm{Ext}^n(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = 0$ for $n \geq 2$. We also have $\mathrm{Ext}^0(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \mathrm{Hom}(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = \{x \in \mathbb{Z}/b\mathbb{Z} : \mathrm{ord}(x)|a\} = (b/d)\cdot\mathbb{Z}/b\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$. And, again from the sequence,

$$\mathrm{Ext}^1(\mathbb{Z}/a\mathbb{Z},\mathbb{Z}/b\mathbb{Z}) = (\mathbb{Z}/b\mathbb{Z})/((a\mathbb{Z}+b\mathbb{Z})/b\mathbb{Z}) \cong \mathbb{Z}/d\mathbb{Z}.$$

□

What about Ext's involving $\mathbb{Q}$? As $\mathbb{Q}$ is injective, for any abelian group A,

$$\mathrm{Ext}^n(A,\mathbb{Q}) = 0, \quad n \geq 1.$$

Determining $\mathrm{Ext}^0(A,\mathbb{Q}) = \mathrm{Hom}(A,\mathbb{Q})$ is interesting; it is easy for a finitely generated abelian group, but complicated in general, and we will not discuss it further here.

What about $\mathrm{Ext}^n(\mathbb{Q},B)$ then? Or even just $\mathrm{Ext}^n(\mathbb{Q},\mathbb{Z})$? Recall the injective resolution of $\mathbb{Z}$:

$$0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0,$$

that provides us (using the functor $\mathrm{Hom}(\mathbb{Q},-)$) with

$$0 \to \mathrm{Hom}(\mathbb{Q},\mathbb{Z}) \to \mathrm{Hom}(\mathbb{Q},\mathbb{Q}) \to \mathrm{Hom}(\mathbb{Q},\mathbb{Q}/\mathbb{Z}) \to \mathrm{Ext}^1(\mathbb{Q},\mathbb{Z}) \to \mathrm{Ext}^1(\mathbb{Q},\mathbb{Q}) = 0.$$

Because $\mathbb{Q}$ is divisible, $\mathrm{Hom}(\mathbb{Q},\mathbb{Z}) = \{0\}$ and because $\mathbb{Q}$ is uniquely divisible, you can show that $\mathrm{Hom}(\mathbb{Q},\mathbb{Q}) \cong \mathbb{Q}$ by $f \mapsto f(1)$ (the subtle point is that we are talking about maps of $\mathbb{Q}$ to $\mathbb{Q}$ as abelian groups, so something needs to be proven). Thus, we have an exact sequence

$$0 \to \mathbb{Q} \to \mathrm{Hom}(\mathbb{Q},\mathbb{Q}/\mathbb{Z}) \to \mathrm{Ext}^1(\mathbb{Q},\mathbb{Z}) \to 0.$$

The first arrow takes an element $t \in \mathbb{Q}$ to the map $x \mapsto xt \pmod{\mathbb{Z}}$. It turns out that $\mathrm{Hom}(\mathbb{Q},\mathbb{Q}/\mathbb{Z})$ is much more complicated. A rather surprising theorem of J. Wiegold, Bull. Austral. Math. Soc., vol. 1 (1969), 341-343, states that

$$\mathrm{Ext}^1(\mathbb{Q},\mathbb{Z}) \cong \mathbb{R}.$$

(Although the isomorphism is "abstract"; it is not a natural isomorphism.)

38. Remarks about the geometric situation

Let $(X,\mathcal{O}_X)$ be a connected compact complex manifold and consider the exact sequence of sheaves on X:

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathcal{O}_X \xrightarrow{e^{2\pi i(-)}} \mathcal{O}_X^\times \longrightarrow 0.$$

Applying the global sections functor, and using the maximum principle, we get,

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{C} \longrightarrow \mathbb{C}^\times \longrightarrow H^1(X,\mathbb{Z}) \longrightarrow H^1(X,\mathcal{O}_X) \longrightarrow H^1(X,\mathcal{O}_X^\times) \longrightarrow H^2(X,\mathbb{Z}) \longrightarrow \dots$$

From which we extract

$$0 \longrightarrow H^1(X,\mathbb{Z}) \longrightarrow H^1(X,\mathcal{O}_X) \longrightarrow H^1(X,\mathcal{O}_X^\times) \longrightarrow H^2(X,\mathbb{Z}) \longrightarrow \dots$$

We have the following interesting facts:

(1) $H^i(X,\mathbb{Z}) \cong H^i_{\mathrm{Betti}}(X,\mathbb{Z})$.

(2) $H^1(X, \mathcal{O}_X^\times) \cong \mathrm{Pic}(X)$ classifies isomorphism classes of line bundles on X.
(3) If X is n-dimensional, $H^q(X, \mathcal{O}_X) = 0$ for $q > n$.

So, for example, if X is a Riemann surface of genus g we find

$$\begin{array}{ccccccccc} 0 & \longrightarrow & H^1(X,\mathbb{Z}) & \longrightarrow & H^1(X,\mathcal{O}_X) & \longrightarrow & H^1(X,\mathcal{O}_X^\times) & \xrightarrow{\delta} & H^2(X,\mathbb{Z}) & \longrightarrow & 0 \\ & & \| & & & & & & \| & & \\ & & \mathbb{Z}^{2g} & & & & & & \mathbb{Z} & & \end{array}$$

It is a fact that $H^1(X, \mathcal{O}_X)$ is a complex vector space of dimension g, and that $H^1(X, \mathbb{Z})$ embeds as a discrete subgroup in it. One deduces that

$$\mathrm{Pic}^0(X) := \mathrm{Ker}(\delta) \cong H^1(X, \mathcal{O}_X)/H^1(X, \mathbb{Z})$$

is a complex torus of dimension g (topologically, a product of $2g$ circles). This is called the **Jacobian** of X. The map $H^1(X, \mathcal{O}_X^\times) \to H^2(X, \mathbb{Z}) \cong \mathbb{Z}$, is the map associating to a line bundle its degree.

39. Ext^1 and extensions

39.1. **Extensions of R-modules.** An extension ξ of R-modules is an exact sequence of R-modules

$$\xi \qquad 0 \longrightarrow A \xrightarrow{f} E \xrightarrow{g} C \longrightarrow 0.$$

It is called an **extension of A by C**; it will be related to $\mathrm{Ext}^1(C, A)$. The extension is called **split** if exists a homomorphism $h\colon C \to E$ such that $gh = Id_C$. It implies that $E \cong A \oplus C$. In fact, $E = f(A) \oplus h(C)$. Such a homomorphism h is called a **section** of the homomorphism g.

Conversely, if we can find a decomposition $E \cong A \oplus C'$ such that f corresponds to the inclusion of A in the first coordinate and g induces an isomorphism $\{0\} \oplus C' \to C$, then the extension splits.

Two extensions ξ, ξ' of A by C are called **equivalent** if there is a commutative diagram

$$\begin{array}{lccccccc} \xi & 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C & \longrightarrow & 0 \\ & & & \| & & \downarrow{\scriptstyle \varphi} & & \| & & \\ \xi' & 0 & \longrightarrow & A & \longrightarrow & E' & \longrightarrow & C & \longrightarrow & 0. \end{array}$$

Note that φ is automatically an isomorphism. It is possible that there is an isomorphism $E \cong E'$, but there is no such isomorphism that is the identity on A and induced the identity on C. The goal is to classify extensions of A by C up to equivalence. As an example, the following extensions are not equivalent:

$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0\,, \qquad 0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow \mathbb{Z}/4\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

They are not equivalent simply because $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \ncong \mathbb{Z}/4\mathbb{Z}$, but in general there exist more subtle examples, examples where $E \cong E'$ as R-modules, but there is no such isomorphism $E \cong E'$ inducing the identity of A and C.

As another example, we prove the following.

Lemma 39.1.1. *The equivalence class of the extension $0 \to A \to A \oplus C \to C \to 0$ is precisely the split extensions.*

Proof. In the notation above, if the extension ξ is equivalent to $0 \to A \to A \oplus C \to C \to 0$, we have a commutative diagram:

$$\begin{array}{ccccccccc} & & 0 & \longrightarrow & A & \longrightarrow & A \oplus C & \longrightarrow & C & \longrightarrow & 0 \\ & & & & \| & & \downarrow{\scriptstyle \varphi} & & \| & & \\ \xi & & 0 & \longrightarrow & A & \xrightarrow{f} & E & \xrightarrow{g} & C & \longrightarrow & 0. \end{array}$$

To simplify notation we view C as a submodule of $A \oplus C$ via $c \mapsto (0, c)$. The diagram implies that $g\colon \varphi(C) \to C$ is an isomorphism. That is, $\varphi\colon C \to E$ is a section of g.

Conversely, suppose that the extension ξ splits; there is a homomorphism $\varphi'\colon C \to E$ splitting the projection $g\colon E \to C$. Meaning $E = f(A) \oplus \varphi'(C)$. Define then $\varphi\colon A \oplus C \to E$ by

$$\varphi(a, c) = f(a) + \varphi'(c).$$

One easily verifies that this provides an isomorphism $A \oplus C \to E$ that is an isomorphism of extensions. □

39.2. **$\mathbf{Ext}^1(C, A) = e(C, A)$.** Our final result would be that equivalence classes of extensions of A by C are in bijection with $\mathrm{Ext}^1(C, A)$. Let $e(C, A)$ denote the equivalence classes of extensions of A by C. We construct an isomorphism

$$\Psi : e(C, A) \to \mathrm{Ext}^1(C, A).$$

To define the map Ψ take a projective resolution $P_\bullet \to C \to 0$ of C and, using the Comparison Lemma, extend the identity morphism 1_C:

$$\begin{array}{ccccccccccc} & \cdots & & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow 0 \\ & & & \vdots & & \vdots{\scriptstyle \alpha} & & \vdots & & \downarrow{\scriptstyle 1_C} & \\ \xi & & & 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C & \longrightarrow 0 \end{array} \tag{39}$$

Recall that

$$\mathrm{Ext}^1(C, A) = \frac{\mathrm{Ker}(d_2^* : \mathrm{Hom}(P_1, A) \to \mathrm{Hom}(P_2, A))}{\mathrm{Im}(d_1^* : \mathrm{Hom}(P_0, A) \to \mathrm{Hom}(P_1, A))}$$

Now, $\alpha \in \mathrm{Ker}\, d_2^*$ because $d_2^*(\alpha) = \alpha \circ d_2 = 0 \circ 0 = 0$. Thus, α provides us with a class $\Psi(\xi)$ in $\mathrm{Ext}^1(C, A)$.

Lemma 39.2.1. *The following holds:*

(1) $\Psi(\xi)$ is independent of the choice of α.
(2) $\Psi(\xi)$ is independent of the choice of resolution.
(3) If $\xi \sim \xi'$ then $\Psi(\xi) = \Psi(\xi')$.

Proof. We start with (1). If we have another extension of 1_C as in (39) giving a map $\alpha'\colon P_1 \to A$ then, by the Comparison Lemma the two extensions of 1_C are homotopic; meaning, there are homomorphisms $s_0\colon P_0 \to A, s_1\colon P_1 \to 0$ etc., such that

$$\alpha - \alpha' = 0 \circ s_1 + s_0 \circ d_1 = d_1^*(s_0).$$

That proves that α and α' are the same element in $\mathrm{Ext}^1(C, A)$.

If we have another resolution, say $P'_\bullet \to C \to 0$, then, using the Comparison Lemma again, we can form a comuttative diagram

$$\begin{array}{lccccccccc}
\cdots & P'_2 & \xrightarrow{d'_2} & P'_1 & \xrightarrow{d'_1} & P'_0 & \longrightarrow & C & \longrightarrow & 0 \\
 & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle 1_C} & & \\
\cdots & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\
 & \downarrow & & \downarrow{\scriptstyle \alpha} & & \downarrow & & \downarrow{\scriptstyle 1_C} & & \\
\xi & 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C & \longrightarrow & 0
\end{array}$$

We can take $\alpha \circ f_1$ as providing us with the element in $\mathrm{Ext}^1(C, A)$ computed using the resolution $P'_\bullet \to C \to 0$. However, it is precisely the map f_1^* that provides the isomorphism between $\mathrm{Ext}^1(C, A)$ calculated using the resolution $P'_\bullet \to C \to 0$ and calculated using the resolution $P_\bullet \to C \to 0$. That is, under that isomorphism we get the same element in $\mathrm{Ext}^1(C, A)$.

To get the last statement, suppose that $\xi \sim \xi'$ and so we have a commutative diagram

$$\begin{array}{lccccccccc}
\cdots & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\
 & \downarrow & & \downarrow{\scriptstyle \alpha} & & \downarrow & & \downarrow{\scriptstyle 1_C} & & \\
\xi & 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C & \longrightarrow & 0 \\
 & & & \| & & \downarrow{\scriptstyle \varphi} & & \| & & \\
\xi' & 0 & \longrightarrow & A & \longrightarrow & E' & \longrightarrow & C & \longrightarrow & 0.
\end{array}$$

But then, directly from the definition, we get the same element in $\mathrm{Ext}^1(C, A)$.

□

Remark 39.2.2. There is another way to get the element $\Psi(\xi)$. Apply $\mathrm{Hom}(-, A)$ to the exact sequence $0 \to A \to E \to C \to 0$ to get

$$0 \to \mathrm{Hom}(C, A) \to \mathrm{Hom}(E, A) \to \mathrm{Hom}(A, A) \to \mathrm{Ext}^1(C, A).$$

The image of the identity map 1_A under the map $\mathrm{Hom}(A, A) \to \mathrm{Ext}^1(C, A)$ provides us with an element of the latter. It is true, but exhausting to prove, that this element is precisely $\Psi(\xi)$. The advantage, though, once you proved this comparison, that it is clearly independent of all choices!

We now construct an inverse Θ to Ψ:

$$\Theta\colon \mathrm{Ext}^1(C, A) \to e(C, A).$$

Start with a projective resolution of C,

$$\cdots \longrightarrow P_2 \xrightarrow{d_2} P_1 \xrightarrow{d_1} P_0 \longrightarrow C \longrightarrow 0$$

An element of $\mathrm{Ext}^1(C, A)$ is an element $\alpha \in \mathrm{Hom}(P_1, A)$ such that $d_2^*(\alpha) = \alpha \circ d_2 = 0$, and α is taken modulo $\mathrm{Im}(d_1^*)$. As $\alpha \circ d_2 = 0$, α induces a diagram as follows

$$\begin{array}{ccccccccc}
0 & \longrightarrow & P_1/\mathrm{Im}(d_2) & \xrightarrow{\bar{d}_1} & P_0 & \xrightarrow{d_0} & C & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \bar{\alpha}} & & & & \downarrow{\scriptstyle 1_C} & & \\
0 & \longrightarrow & A & & & & C & \longrightarrow & 0
\end{array}$$

We then construct E as the push-out:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & P_1/\mathrm{Im}(d_2) & \xrightarrow{\bar{d}_1} & P_0 & \xrightarrow{d_0} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \bar{\alpha}} & & \downarrow & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \longrightarrow & E & & C & \longrightarrow & 0 \end{array}$$

and we claim that there is a natural arrow $E \to C$ making the whole diagram commutative and the bottom row exact:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & P_1/\mathrm{Im}(d_2) & \xrightarrow{\bar{d}_1} & P_0 & \xrightarrow{d_0} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \bar{\alpha}} & & \downarrow & & \downarrow{\scriptstyle 1_C} & & \\ 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C & \longrightarrow & 0 \end{array}$$

The arrow $E \to C$ comes from the push-out property with regard to the morphisms $d_0 \colon P_0 \to C$ and the zero map $A \to C$. Note that these morphisms agree when composed with the morphisms $P_1/\mathrm{Im}(d_2) \to P_0$ and $P_1/\mathrm{Im}(d_2) \to A$ (both compositions being 0) and so the push-out property provides the map $E \to C$.

Lemma 39.2.3. *The lower row in exact.*

Proof. Recall that we have an explicit description of the push out:

$$E = A \oplus P_0/\{(\bar{\alpha}(x), -d_1(x)) : x \in P_1/\mathrm{Im}(d_2)\}.$$

Then, the map $A \to E$ is just $a \mapsto (a,0)$, which is injective because if $(a,0) = (\alpha(x),0)$ then $d_1(x) = 0$, but the map $P_1/\mathrm{Im}(d_2) \to P_0$ is injective so $x = 0$ thus $a = 0$. The map $E \to C$ is just $(a,p) \mapsto d_0(p)$, which is well-defined and surjective because $P_0 \to C$ is surjective. Clearly the composition of the maps $A \to E \to C$ is zero and so the only thing left to check is that any element (a,p) such that $d_0(p) = 0$ is equal in E to an element of the form $(a,0)$. Indeed, as $d_0(p) = 0$, there is some $x \in P_1/\mathrm{Im}(d_2)$ such that $p = d_1(x)$ and then, in E, $(a,p) = (a,p) + (\alpha(x), -d_1(x)) = (a + \alpha(x), 0)$ and it belongs to the image of A in E. □

Thus, we got an extension $\Theta(\alpha)$ associated to an element $\alpha \in \mathrm{Ext}^1(C,A)$. Is it well-defined? The point is that α is really a coset and not a single homomorphism. That is, we need to show the following lemma:

Lemma 39.2.4. *Let $\alpha' \in \alpha + \mathrm{Im}(d_1^* \colon \mathrm{Hom}(P_0, A) \to \mathrm{Hom}(P_1, A))$ be another element in the coset of α, then α' defines an equivalent extension.*

Proof. Such an element α' has the form

$$\alpha' = \alpha + s \circ d_1$$

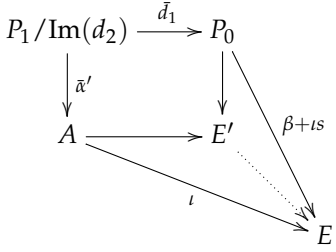
for some homomorphism $s \colon P_0 \to A$. The first push-out construction gave us

$$\begin{array}{ccc} P_1/\mathrm{Im}(d_2) & \xrightarrow{\bar{d}_1} & P_0 \\ \downarrow{\scriptstyle \bar{\alpha}} & & \downarrow{\scriptstyle \beta} \\ A & \xrightarrow{\iota} & E \end{array} \quad , \qquad \iota\bar{\alpha} = \beta\bar{d}_1$$

Then,

$$\iota(\bar{\alpha} + s\bar{d}_1) = \iota\bar{\alpha} + \iota s\bar{d}_1 = (\beta + \iota s)\bar{d}_1.$$

This means that if we let E' the pushout for $(\bar{\alpha}', \bar{d_1})$ then, using the push-out property of E', we get a commutative diagram



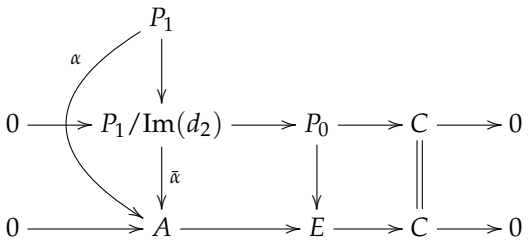
Recalling how the maps $E \to C$ and $E' \to C$ were constructed, we see that the morphism $E' \to E$ is a morphism of extensions of A by C and is thus, automatically an isomorphism. □

Theorem 39.2.5. *There is an isomorphism*

$$\mathrm{Ext}^1(C, A) \cong e(C, A).$$

Proof. Having constructed Ψ and Θ what remains is to show $\Theta \circ \Psi$ and $\Psi \circ \Theta$ are the identity.

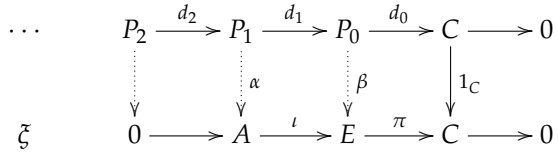
$\boxed{\Psi\Theta = \mathrm{Id}}$ Θ starts with a projective resolution $P_\bullet \to C \to 0$ and an $\alpha\colon P_1 \to A$. We refer to the diagram



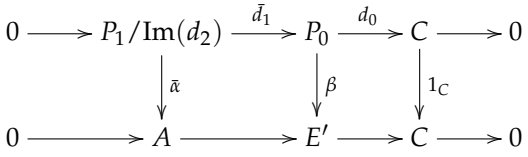
It induces then the diagram (39) and that shows that $\Psi\Theta\alpha$ is just the class of α in $\mathrm{Ext}^1(C, A)$.

$\boxed{\Theta\Psi = \mathrm{Id}}$ Here we start with the diagram (39)

(40)



Then we have argued that $\alpha \in \mathrm{Ker}(d_2^*)$ and that gave us a class in $\mathrm{Ext}^1(C, A)$. But α also induces $\bar{\alpha}\colon P_1/\mathrm{Im}(d_2) \to A$ and then the procedure for calculation Θ is to form the commutative diagram where E' is defined as a push-out



We need to show that $E \cong E'$ as extensions. We claim that E is a push-out. The Snake Lemma given that $\mathrm{Coker}(\bar{\alpha}) \cong \mathrm{Coker}(\beta)$ and that implies that the map

$$A \oplus P_0 \to E, \quad (a, p) \mapsto \iota(a) + \beta(p)$$

is surjective. Suppose that (a, p) is in kernel, meaning $-\iota(a) = \beta(p)$. Then $d_0(p) = \pi\beta(p) = \pi\iota(a) = 0$ and so $p = \bar{d_1}(x)$ for some unique $x \in P_1/\mathrm{Im}(d_2)$. Now, $\iota(\bar{\alpha}(x)) = \beta\bar{d_1}(x) = \beta(p) = -\iota(a)$ and, since ι is injective, we must have $\bar{\alpha}(-x) = a$ and the element in the kernel has the form $(\alpha(-x), -d_1(-x))$. Our explicit expression for the push-out implies that E is the push-out. □

Example 39.2.6. As $\mathrm{Ext}^1(C,A)$ is an abelian group one can ask how the group law is phrased in terms of extensions. We discuss it in § 39.3 below. An easier question is how to describe the inverse of an extension ξ under the group law. Consider an extension

$$\xi \qquad 0 \longrightarrow A \xrightarrow{f} E \xrightarrow{g} C \longrightarrow 0.$$

We claim that $-\xi$ is given by the extension

$$-\xi \qquad 0 \longrightarrow A \xrightarrow{-f} E \xrightarrow{g} C \longrightarrow 0.$$

To see that recall the definition of the isomorphism $\Psi\colon e(C,A) \to \mathrm{Ext}^1(C,A)$. To the extension ξ we associated the element $\alpha \in \mathrm{Ext}^1(C,A)$, where α came through the Comparison Lemma:

$$\begin{array}{lccccccccc}
\cdots & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\
 & \downarrow & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle s} & & \downarrow{\scriptstyle 1_C} & & \\
\xi & 0 & \longrightarrow & A & \xrightarrow{f} & E & \xrightarrow{g} & C & \longrightarrow & 0
\end{array}$$

But then we have

$$\begin{array}{lccccccccc}
\cdots & P_2 & \xrightarrow{d_2} & P_1 & \xrightarrow{d_1} & P_0 & \longrightarrow & C & \longrightarrow & 0 \\
 & \downarrow & & \downarrow{\scriptstyle -\alpha} & & \downarrow{\scriptstyle s} & & \downarrow{\scriptstyle 1_C} & & \\
-\xi & 0 & \longrightarrow & A & \xrightarrow{-f} & E & \xrightarrow{g} & C & \longrightarrow & 0
\end{array}$$

and that proves our assertion, as $-\alpha$ is clearly the inverse of α in $\mathrm{Ext}^1(C,A)$.

Note that multiplication by $[-1]$ induces a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A & \xrightarrow{-f} & E & \longrightarrow & C & \xrightarrow{g} & 0 \\
 & & \| & & \downarrow{\scriptstyle [-1]} & & \| & & \\
0 & \longrightarrow & A & \xrightarrow{f} & E' & \xrightarrow{-g} & C & \longrightarrow & 0.
\end{array}$$

Thus, we could have equally claimed that $-\xi$ is given by

$$0 \longrightarrow A \xrightarrow{f} E' \xrightarrow{-g} C \longrightarrow 0.$$

39.3. **Baer sum.** Since $\mathrm{Ext}^1(C,A)$ is an abelian group, this implies that there should be a way to *add* two extensions of A by C. The **Baer sum** is this addition operation. We describe it, but omit the proofs. They can be found in Rotman's book and are best done as an exercise.

Pushout and pullback of extensions. Given an extension $\xi\colon\ 0 \to A \to E \to C \to 0$ and a homomorphism $h\colon A \to A'$ show that there is a diagram

$$\begin{array}{lccccccccc}
\xi & 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C & \longrightarrow & 0 \\
 & & & \downarrow{\scriptstyle h} & & \downarrow & & \| & & \\
h_*(\xi) & 0 & \longrightarrow & A' & \longrightarrow & E' & \longrightarrow & C & \longrightarrow & 0,
\end{array}$$

where one first constructs E' as a push-out and argues that there is a natural map $E' \to C$ making the diagram commutative, and E' an extension of A' by C. This is in fact the functorial map

$$h_*\colon \mathrm{Ext}^1(C,A) \to \mathrm{Ext}^1(C,A'),$$

coming from the formalism of derived functors ($\mathrm{Ext}^1(C,A)$ is an additive covariant functor in A), but the proof is laborious and one might want to look up the proof.

Similarly, given an extension $\xi\colon\ 0 \to A \to E \to C \to 0$ and a homomorphism $k\colon C' \to C$ show that there is a diagram

$$\begin{array}{lccccccc} k^*(\xi) & 0 & \longrightarrow & A & \dashrightarrow & E' & \dashrightarrow & C' \longrightarrow 0 \\ & & & \| & & \downarrow & & \downarrow k \\ \xi & 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & C \longrightarrow 0, \end{array}$$

where one first constructs E' as a pull-back and argues that there is a natural map $A \to E'$ making the diagram commutative and E' an extension of A by C'. This is the functorial map

$$k^*\colon \mathrm{Ext}^1(C,A) \to \mathrm{Ext}^1(C',A),$$

coming from the formalism of derived functors ($\mathrm{Ext}^1(C,A)$ is an additive contravariant functor in C); again, the proof is laborious and one might want to look it up.

The Baer sum of two extensions. Let $\xi_i\colon\ 0 \to A \to E_i \to C \to 0$, $i = 1,2$ be extensions of A by C. Then, we have an extension

$$\xi_1 \oplus \xi_2\colon\quad 0 \to A \oplus A \to E_1 \oplus E_2 \to C \oplus C \to 0, \qquad \xi_1 \oplus \xi_2 \in \mathrm{Ext}^1(C \oplus C, A \oplus A).$$

Let Δ_C be the diagonal homomorphism,

$$\Delta_C\colon C \to C \oplus C, \quad \Delta(c) = (c,c).$$

We then have the class $\Delta_C^*(\xi_1 \oplus \xi_2) \in \mathrm{Ext}^1(C, A \oplus A)$. We also have the sum homomorphism

$$\nabla_A\colon A \oplus A \to A, \quad \nabla_A(a,a') = a + a'.$$

Using it we get a class $\nabla_{A,*}\Delta_C^*(\xi_1 \oplus \xi_2) \in \mathrm{Ext}^1(C,A)$. This the Baer sum:

$$\xi_1 + \xi_2 = \nabla_{A,*}\Delta_C^*(\xi_1 \oplus \xi_2)$$

40. Group cohomology

Let G be a group and consider $\mathbb{Z}[G]$-modules; equivalently, abelian groups on which G acts by group automorphisms from the left. That is, we consider the category ${}_{\mathbb{Z}[G]}\mathbf{Mod}$. The group G itself need not be abelian (or finite). So this is at the outset quite a general situation. However, we will quickly restrict our discussion to finite groups G.

For a $\mathbb{Z}[G]$-module A, let

$$A^G = \{a \in A : ga = a, \forall g \in G\} = \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}, A)$$

where $\mathbb{Z}$ is given the trivial G-action. Thus, the left-exact covariant additive functor $A \mapsto A^G$ can be identified with $\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}, -)$ and we denote its right derived-functors

$$H^n(G, A) = \mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z}, A)$$

We added the subscript $\mathbb{Z}[G]$ to stress that the derived functors are calculated in the category of $\mathbb{Z}[G]$-modules and not in the category of $\mathbb{Z}$-modules. Thus, we can rephrase a result, which we proved in general, as follows: given a short exact sequence of $\mathbb{Z}[G]$-modules

$$0 \to A \to B \to C \to 0,$$

there is a long exact sequence of abelian groups

$$0 \to A^G \to B^G \to C^G \to H^1(G, A) \to H^1(G, B) \to H^1(G, C) \to H^2(G, A) \to \cdots$$

It should be remarked that although the cohomology groups $H^n(G, A)$ were introduced to study the failure of the fixed points functor to be exact, they are interesting invariants of the group in themselves, and have other uses, some of which will be mentioned below.

Here are some examples that we will return to later.

(1) Let L/K be a finite Galois extension with Galois group $G = \mathrm{Gal}(L/K)$. Then both $L^\times$ with multiplication, and L with addition, are abelian groups with G-action. The invariants are of course just $K^\times$ and K, respectively, but the group cohomology of G with values in $L^\times$, for example, is important.

(2) Let X be a complex manifold and $G \subseteq \mathrm{Aut}(X)$ a group of automorphisms of X. Let M be the field of meromorphic functions on X (globally defined). Then g acts on $M^\times = M - \{0\}$ by

$$({}^g f)(x) := f(g^{-1}x).$$

If G is finite and acts freely, orbit space $Y := X/G$ can be given a structure of a complex manifold such that $X \to Y$ is a holomorphic covering map with deck transformations G. The invariants M^G are precisely the meromorphic functions on Y. The norm map and trace maps

$$M \to M^G, \quad f \mapsto \mathrm{Nm}(f) := \prod_{g \in G} {}^g f, \quad f \mapsto \mathrm{Tr}(f) := \sum_{g \in G} {}^g f,$$

are interesting from a geometric point of view. For example, the kernel of the trace map are the functions that average over every orbit to 0. How can we find such functions? Once again, group cohomology enters the picture.

(3) Fix a group G and suppose that an abelian group A arises in an exact sequence of groups

$$1 \to A \to H \to G \to 1.$$

Then G acts on A as follows: given an element g of G, lift it to an element $\tilde{g} \in H$ and define for $a \in A$

$${}^g a := \tilde{g} a \tilde{g}^{-1}.$$

The element ${}^g a$ doesn't depend on the choice of lift $\tilde{g}$ because any two choices differ by an element of A and A is abelian. Galois cohomology then applies to the classification of such extensions.

40.1. **The standard resolution $Q_\bullet$ of $\mathbb{Z}$ by $\mathbb{Z}[G]$-modules.** Recall that to calculate $H^n(G,A)$, that is, to calculate $\mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z},A)$, we may either use an injective resolution of A in $\mathbb{Z}[G]$-modules, or a projective resolution of $\mathbb{Z}$ in $\mathbb{Z}[G]$-modules. Projective modules are easier to handle than injective ones and so we use the latter option.

We will assume henceforth that G is a finite group. For $n>0$, let Q_n be the free $\mathbb{Z}[G]$-module on the basis

$$\{[x_1,\dots,x_n] : x_i \in G\}.$$

We also define

$$Q_0 = \mathbb{Z}[G]\cdot[\,],$$

where $[\,]$ is a formal symbol introduced in order to streamline notation. Thus, Q_0 is a free $\mathbb{Z}[G]$-module of rank 1; more generally, Q_n is a free $\mathbb{Z}[G]$-module of rank $\sharp G^n$.

Let

$$\epsilon\colon Q_0 \to \mathbb{Z}, \quad \epsilon(\sum_{g\in G} a_g g\cdot[\,]) = \sum_{g\in G} a_g.$$

Define for $n\geq 1$,

$$d_n\colon \mathbb{Q}_n \to \mathbb{Q}_{n-1},$$

by defining d_n on the basis elements and extending the definition $\mathbb{Z}[G]$-linearly:

$$d_n([x_1,\dots,x_n]) = x_1[x_2,\dots,x_n] + \sum_{i=1}^{n-1}(-1)^i[x_1,\dots,x_ix_{i+1},\dots,x_n] + (-1)^n[x_1,\dots,x_{n-1}].$$

For example,

- $d_1[x_1] = x_1\cdot[\,]-[\,]$.
- $d_2[x_1,x_2] = x_1[x_2]-[x_1x_2]+[x_1]$.
- $d_3[x_1,x_2,x_3] = x_1[x_2,x_3]-[x_1x_2,x_3]+[x_1,x_2x_3]-[x_1,x_2]$.

Lemma 40.1.1. *The construction above provides a projective resolution of $\mathbb{Z}$ as a $\mathbb{Z}[G]$-module:*

$$\cdots \longrightarrow Q_2 \xrightarrow{d_2} Q_1 \xrightarrow{d_1} Q_0 \xrightarrow{\epsilon} \mathbb{Z} \longrightarrow 0.$$

*The elements in the kernel of $Q_n \to Q_{n-1}$ are called n-***cochains** *and those in the image of $Q_{n+1}\to Q_n$ are called n-***coboundaries**.

Proof. We show exactness at Q_0 and a bit more. The complete proof is actually somewhat complicated (you can try to prove exactness at Q_1 to get some appreciation for it). It can be found in Rotman's book, §10.

It is clear that ϵ is surjective. Its kernel is $\{\sum_g a_g g\cdot[\,] : a_g\in\mathbb{Z}, \sum_g a_g = 0\}$. The kernel of ϵ is a $\mathbb{Z}[G]$-module, thus, to show that $\mathrm{Im}(d_1)\subseteq \mathrm{Ker}(\epsilon)$ it is enough to examine the generators, namely, to show that $\epsilon(d_1[x_1])=0$. But, $d_1[x_1] = x_1\cdot[\,]-[\,]$. To be more pedantic it is $1\cdot x_1[\,]-1\cdot 1_G[\,]$ and it maps to 0 under ϵ.

Now, suppose that $\sum_g a_g g[\,]\in\mathrm{Ker}(\epsilon)$ then $(\sum_g a_g)[\,]=0$ and therefore

$$\sum_g a_g g[\,] = \sum_g a_g(g[\,]-[\,]) = d_1(\sum_g a_g[g])$$

is in the image of d_1. It follows that the sequence is exact at Q_0.

The computation that we get a complex is a computation similar to one sees in many contexts, for example in the context of homology in topology. It suffices to prove that $d_{n-1}d_n[x_1,\dots,x_n]=0$.

$$d_{n-1}d_n([x_1,\dots,x_n])$$
$$= x_1 d_{n-1}[x_2,\dots,x_n] + \sum_{i=1}^{n-1}(-1)^i d_{n-1}[x_1,\dots,x_ix_{i+1},\dots,x_n] + (-1)^n d_{n-1}[x_1,\dots,x_{n-1}].$$

This is complicated to prove in one shot, so we look at this expression according to the type of elements one finds when writing it down explicitly. When expanding this expression, there are terms of the form

$$x_1x_2[\cdots].$$

They can only come from the first summand, or from the first summand in the Σ and we get

$$x_1x_2[x_3,\dots,x_n] + (-1)^1 x_1x_2[x_3,\dots,x_n] = 0.$$

There are also terms of the form

$$x_1[\dots].$$

The ones arising from the first summand are $(-1)^{i-1}x_1[x_2,\dots,x_ix_{i+1},\dots,x_n]$ and they cancel with the first term of $(-1)^i d_{n-1}[x_1,\dots,x_ix_{i+1},\dots,x_n]$, except for the term $(-1)^{n-1}x_1[x_2,\dots,x_{n-1}]$ that cancels with the first term in $(-1)^n d_{n-1}[x_1,\dots,x_{n-1}]$.

Then there are terms of the form

$$(-1)^n(-1)^i[x_1,\dots,x_ix_{i+1},\dots,x_{n-1}]$$

coming from the last summand. Those are cancelled by the last term of (the i-the term of the Σ) $(-1)^i d_{n-1}[x_1,\dots,x_ix_{i+1},\dots,x_n]$.

There are also the terms

$$[x_1,\dots,x_ix_{i+1},\dots,x_jx_{j+1},\dots,x_n].$$

These arise from either (the i-th term of the Σ) $(-1)^i d_{n-1}[x_1,\dots,x_ix_{i+1},\dots,x_n]$ appearing with sign $(-1)^i(-1)^{j-1}$, or from (the j-th term of the Σ) $(-1)^j d_{n-1}[x_1,\dots,x_jx_{j+1},\dots,x_n]$ appearing with sign $(-1)^j(-1)^i$, hence cancel.

Finally, there are terms of the form $[x_1,\dots,x_{i-1}x_ix_{i+1},\dots,x_n]$ that arise from the $i-1$-st and i-th terms of the Σ. That is, from expanding $(-1)^{i-1}d_{n-1}[x_1,\dots,x_{i-1}x_i,\dots,x_n]$ and from expanding $(-1)^i d_{n-1}[x_1,\dots,x_ix_{i+1},\dots,x_n]$, and they have opposite signs hence cancel.

We have dealt this way with all the possible types of terms, thus $d_{n-1}d_n = 0$. It is perhaps a good idea to write this out explicitly for d_2d_3 to understand what's going on. □

Using this resolution we can find a concrete description of $H^n(G,A)$. Note that

$$\mathrm{Hom}_{\mathbb{Z}[G]}(Q_n,A) \cong \{\text{functions } f\colon G^n \to A\},$$

because a $\mathbb{Z}[G]$-homomorphism from Q_n to A is uniquely determined by its value on the basis elements $\{[x_1,\dots,x_n]\}$ and, conversely, assigning any values in A for the basis elements provide a function that extends to a $\mathbb{Z}[G]$-homomorphism $Q_n \to A$. Recall also that

$$H^n(G,A) = \mathrm{Ker}(d_{n+1}^*)/\mathrm{Im}(d_n^*).$$

Analyzing $\mathrm{Ker}(d_{n+1}^*)$ and $\mathrm{Im}(d_n^*)$, we find, for example:

- $H^0(G,A) = \{a \in A : ga - a = 0, \forall g \in G\} = A^G$.

- $H^1(G,A) = \frac{\{f:G\to A \mid x_1 f(x_2) - f(x_1x_2) + f(x_1) = 0\}}{\{f:G\to A \mid f(x) = xa - a \text{ for some } a\in A, \text{ independent of } x\}}$.
- $H^2(G,A) = \frac{\{f:G\times G\to A \mid x_1 f(x_2,x_3) - f(x_1x_2,x_3) + f(x_1,x_2x_3) - f(x_1,x_2) = 0\}}{\{f:G\times G\to A \mid f(x_1,x_2) = x_1 g(x_2) - g(x_1x_2) + g(x_1), \text{ for some } g:G\to A\}}$.

Very rarely one uses the explicit description of $H^n(G,A)$ for $n > 2$. So these formulas are by far the most important cases.

Remark 40.1.2. Note that *if G acts trivially on A*,

$$H^1(G,A) = \mathrm{Hom}(G,A),$$

where $\mathrm{Hom}(G,A)$ means the group homomorphisms $G \to A$. This is very useful, and often occurs, in applications.

40.2. **Hilbert's Theorem 90 and Kummer's theory.**

Theorem 40.2.1 (Hilbert's 90)**.** *Let L/K be a finite Galois extension with Galois group G. Then*

$$H^1(G, L^\times) = 0.$$

Proof. Any class in $H^1(G,L^\times)$ is represented by a function $\varphi\colon G \to L^\times$, such that

$$\varphi(\tau\sigma) = {}^\tau\varphi(\sigma)\cdot\varphi(\tau).$$

(This is the cocycle condition when the group law on the abelian group is written multiplicatively.) To show that φ is the trivial element of $H^1(G,L^\times)$ is to show that there is an element $c \in L^\times$ such that

$$\varphi(\tau) = \tau(c)/c, \quad \forall \tau \in G.$$

The functions $\sigma\colon L^\times \to L^\times$ are characters. By *Independence of Characters* any non-trivial linear combination of them is not identically zero. In particular, taking as coefficients $\varphi(\sigma)$ we find that the function from $L^\times$ to $L^\times$ given by

$$\sum_{\sigma\in G} \varphi(\sigma)\sigma(\cdot) \neq 0.$$

That is, there is an element $a \in L^\times$ such that

$$b := \sum_{\sigma\in G} \varphi(\sigma)\sigma(a) \neq 0.$$

Now,

$${}^\tau b = \sum_{\sigma\in G} {}^\tau\varphi(\sigma)\tau\sigma(a) = \varphi(\tau)^{-1}\sum_{\sigma\in G}\varphi(\tau\sigma)\tau\sigma(a) = \varphi(\tau)^{-1}b.$$

Let $c = b^{-1}$. It follows that

$$\varphi(\tau) = b/\tau(b) = \tau(c)/c, \forall \tau \in G.$$

□

Lemma 40.2.2. *In the setting of Hilbert's 90, assume that $\mu_n \subset K$*[56] *then*

$$(K^\times \cap (L^\times)^n)/(K^\times)^n \cong \mathrm{Hom}(G,\mu_n).$$

[56] By that we mean that K contains n distinct n-th roots of unity; they are denoted μ_n.

Proof. We have the exact sequence

$$0 \longrightarrow \mu_n \longrightarrow L^\times \xrightarrow{x\mapsto x^n} L^{\times,n} \longrightarrow 0\,,$$

where $L^{\times,n}$ is the subgroup of $L^\times$ comprised elements of $L^\times$ that are an n-th power of another element in $L^\times$. Taking G-invariants we find

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mu_n & \longrightarrow & K^\times & \xrightarrow{x\mapsto x^n} & K^\times \cap L^{\times,n} & \xrightarrow{\delta} & H^1(G,\mu_n) \longrightarrow H^1(G,L^\times) \\ & & & & & & & & \| \text{trivial action} \qquad \| \text{Hilbert's 90} \\ & & & & & & & & \mathrm{Hom}(G,\mu_n) \qquad 0 \end{array}$$

□

Remark 40.2.3. The isomorphism

$$\delta\colon \frac{K^\times \cap L^{\times,n}}{K^{\times,n}} \longrightarrow \mathrm{Hom}(G,\mu_n)$$

has a simple explicit direction that one can find by recalling how the long exact sequence comes about. We have

$$\delta(\lambda) = \{\sigma \mapsto {}^\sigma t/t\}, \qquad t^n = \lambda$$

The following corollary is the beginning of **Kummer theory** in its modern formulation.

Corollary 40.2.4. *Let L/K be a cyclic Galois extension of order n, such that $\mu_n \subset K$. Then, $(K^\times \cap (L^\times)^n)/(K^\times)^n$ is a cyclic group of order n, generated by α^n for some element $\alpha \in L^\times$ such that $\alpha^n \in K^\times$. Moreover, $L = K(\alpha)$ and the homomorphism corresponding to α^n is*

$$\varphi : G \to \mu_n, \quad \varphi(\sigma) = \sigma(\alpha)/\alpha.$$

Proof. Note that in this case $\mathrm{Hom}(G,\mu_n)$ is a cyclic group of order n. Under the assumptions, there is an isomorphism $h\colon G \to \mu_n$. Thus, there is an element $\lambda \in K^\times \cap L^{\times,n}, \lambda = \alpha^n$, such that $\delta(\lambda) = h$. But then

$$\delta(\lambda)(\sigma) = {}^\sigma\alpha/\alpha = h(\sigma) \neq 1, \quad \forall \sigma \neq 1_G.$$

This proves that $L = K(\alpha)$. □

Further analysis yields one of the main results of Kummer's theory. It is a good examples of what one is after in many classification problems; it solves the particular classification problem related to K, namely "classify the cyclic Galois extensions of K", in terms of data dependent on K alone (the group $K^\times/(K^\times)^n$):

Theorem 40.2.5. *Assume that K contains n-distinct roots of unity. There exists a bijective correspondence between cyclic subgroups of order n of $K^\times/(K^\times)^n$ and cyclic Galois extensions L/K of order n. Under this correspondence, a cyclic subgroup $\langle\alpha\rangle$ is mapped to the cyclic extension $K(\sqrt[n]{\alpha})$, and a cyclic extension L/K to the subgroup $K^\times \cap L^{\times,n}/K^{\times,n}$.*

40.3. **The cohomology of a cyclic group.** The proofs in this section are left as an exercise. Let $G = \{1, g, \dots, g^{n-1}\}$ be a cyclic group.

There is a canonical projective resolution of $\mathbb{Z}$ by $\mathbb{Z}[G]$-modules, much simpler than the one we constructed in the general case, that takes advantage of the assumption that G is cyclic. Let

$$\epsilon : \mathbb{Z}[G] \to \mathbb{Z}, \quad \epsilon(\sum_g a_g g) = \sum_g a_g.$$

Let $\eta = 1 + g + \cdots + g^{n-1} \in \mathbb{Z}[G]$ and define a homomorphism of $\mathbb{Z}[G]$-modules

$$N : \mathbb{Z}[G] \to \mathbb{Z}[G], \quad N(x) = x\eta$$

(multiplying by the element η). Let $\delta = g - 1$ and let

$$D : \mathbb{Z}[G] \to \mathbb{Z}[G], \quad D(x) = x\delta$$

(multiplying by the element δ). Then one can prove that

$$\cdots \longrightarrow \mathbb{Z}[G] \xrightarrow{N} \mathbb{Z}[G] \xrightarrow{D} \mathbb{Z}[G] \xrightarrow{N} \mathbb{Z}[G] \xrightarrow{D} \mathbb{Z}[G] \xrightarrow{\epsilon} \mathbb{Z} \longrightarrow 0$$

is a projective resolution of $\mathbb{Z}$, viewed as G-module with trivial action.

Let A be a $\mathbb{Z}[G]$-module. Using this resolution, one may deduce the following:

$$H^0(G, A) = A^G, \quad H^{2n-1}(G, A) = A[\eta]/\delta A, \quad H^{2n}(G, A) = A^G/\eta A,$$

where $A[\eta] = \{a \in A : \eta a = 0\}$.

An example of particular interest is the following. Let L/K be a cyclic Galois extension of fields with Galois group G. Let $\lambda \in L^\times$ be an element whose norm to $K^\times$ is equal to 1:

$$\prod_{i=0}^{n-1} g^i(\lambda) = 1.$$

Then, one can prove (making use also of Hilbert's 90) that there is an element $\mu \in L^\times$ such that

$$\lambda = g(\mu)/\mu.$$

40.4. **Extensions of groups.** Let A be an abelian group and G an arbitrary group. An **extension** E of A by G is an exact sequence

$$0 \to A \to E \to G \to 1.$$

Note that these are very different extensions than those we considered previously in §39: (i) First, G need not be a commutative group (and to remind ourselves of this fact, we write 1 in $G \to 1$); (2) Second, even if G is commutative, E need not be commutative.

Thus, the situation we consider here is *very different* from the previously studied notion of extensions in a category of modules.

We consider extensions up to equivalence. Two extensions E, E' of A by G are **equivalent** if there is a homomorphism $\varphi\colon E \to E'$ making the diagram commutative (and it implies that φ must be an isomorphism):

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & E & \longrightarrow & G & \longrightarrow & 1 \\ & & \| & & \downarrow{\scriptstyle \varphi} & & \| & & \\ 0 & \longrightarrow & A & \longrightarrow & E' & \longrightarrow & G & \longrightarrow & 1 \end{array}$$

If E is such an extension then it defines an action of G on A: given $g \in G$ and $a \in A$ define

$${}^g a := g * a := \gamma a \gamma^{-1},$$

where γ is an element of E lifting g. One checks that since A is commutative, the action is independent of the choice of lift γ. Thus, an extension makes A into a $\mathbb{Z}[G]$-module. Alternately, we can say that we have a homomorphism

$$\theta\colon G \to \mathrm{Aut}(A).$$

The proof of the following theorem can be found in Rotman, using results from loc. cit. §5 (5.8, 5.9, 5.13) and §10 (10.26, 10.27).

Theorem 40.4.1. *The group cohomology $H^2(G,A)$ classifies equivalence classes of extensions of A by G, with a given action $\theta\colon G \to \mathrm{Aut}(A)$ of G on A. The trivial class corresponds to the extension which is the semi-direct product $A \rtimes_\theta G$.*

In particular, if $H^2(G,A) = \{0\}$ then any extension of A by G is a semi-direct product. Furthermore, if E is a semi-direct product of A by G then the complements of A in E (considered up to conjugation) are classified by $H^1(G,A)$.

To get an idea of what is behind this theorem, given an extension E, choose a set-theoretic section λ (that is, $f \circ \lambda = Id_G$, but λ need not be a group homomorphism)

$$0 \longrightarrow A \longrightarrow E \underset{\lambda}{\overset{f}{\rightleftarrows}} G \longrightarrow 1.$$

Use $+$ for the group law on E (even though it need not be commutative!) and multiplication for the group operation of G and define a function $\varphi\colon G \times G \to A$ by

$$\lambda(x) + \lambda(y) = \varphi(x,y) + \lambda(xy);$$

one finds that

$$^x\varphi(y,z) - \varphi(xy,z) + \varphi(x,yz) - \varphi(x,y) = 0,$$

where $^x a$ denotes the action of $x \in G$ on $a \in A$ (note that some rearrangement of the terms was needed, but this is permissible since A is abelian). This way, we have produced from the extension E an element in $H^2(G,A)$ and that gives some idea how the proof goes.

Lemma 40.4.2.
- *If $\sharp A = n$, then $n \cdot H^i(G,A) = 0$ for all $i \geq 0$.*
- *If $\sharp G = m$, then $m \cdot H^i(G,A) = 0$ for all $i \geq 1$.*

Proof. The first assertion of lemma follows once one checks that a homomorphism $\varphi\colon A \to B$ induces a homomorphism

$$H^i(G,A) \to H^i(G,B),$$

which on the level of co-chains is nothing else then taking a co-chain f to $\varphi \circ f$. In particular, $[n]\colon A \to A$ induces multiplication by n on $H^i(G,A)$. On the other hand, since A is killed by $[n]$ we see that $[n] \circ f = 0$, meaning $H^i(G,A)$ is killed by $[n]$.

The second assertion looks similar but is in fact more subtle. One might argue that a homomorphism of groups $\varphi\colon G \to H$ induces a homomorphism

$$H^i(H,A) \to H^i(G,A)$$

which on the level of co-chains is nothing else then taking a co-chain f to $f \circ \varphi$. This is correct. And it is also correct that if φ is the trivial homomorphism that the map $H^i(H,A) \to H^i(G,A)$ is the zero homomorphism. But when one now applies it to the homomorphism $G \to G, g \mapsto g^m$ this doesn't provide a proof of the second assertion, because it is not clear that this pullback map on cohomology is multiplication by m. (Note that for every integer r we have multiplication by r map on the cohomology – on any abelian group actually, but

only for $r = m$ (or a multiple of m) we can guarantee that the map $g \mapsto g^m$ is a homomorphism of G. This suggest that there is a genuine difficulty here.) Here is the proof:

Let $n \geq 1$ and let $f\colon G^n \to A$ be any function. Define a function $u\colon G^{n-1} \to A$ by the forumla

$$u[g_1, \ldots, g_{n-1}] = \sum_{g \in G} f[g_1, \ldots, g_{n-1}, g].$$

Then we have the following calculation:

$$\sum_{g\in G}(d^*_{n+1}f)[g_1,\ldots,g_n,g] = g_1 \sum_{g\in G} f[g_2,g_3,\ldots,g_n,g] + \sum_{i=1}^{n-1}(-1)^i \sum_{g\in G} f[g_1,\ldots,g_ig_{i+1},\ldots,g_n,g] + (-1)^n \sum_{g\in G} f[g_1,\ldots,g_ng] + (-1)^{n+1}\sum_{g\in G} f[g_1,\ldots,g_n].$$

That is,

$$\begin{aligned}\sum_{g\in G}(d^*_{n+1}f)[g_1,\ldots,g_n,g] &= g_1u[g_2,g_3,\ldots,g_n] + \sum_{i=1}^{n-1}(-1)^i u[g_1,\ldots,g_ig_{i+1},\ldots,g_n] \\ &\quad + (-1)^n u[g_1,\ldots,g_{n-1}] + (-1)^{n+1}\sharp G\cdot f[g_1,\ldots,g_n] \\ &= (d^*_n u)[g_1,\ldots,g_n] + (-1)^{n+1}\sharp G\cdot f[g_1,\ldots,g_n].\end{aligned}$$

Thus, if f defines an element in cohomology, namely if $d^*_{n+1}f = 0$ then we find that

$$\sharp G\cdot f[g_1,\ldots,g_n] = (-1)^{n+1}(d^*_n u)[g_1,\ldots,g_n].$$

Namely, $\sharp G\cdot f[g_1,\ldots,g_n]$ is a coboundary, hence 0 in cohomology. □

Corollary 40.4.3 (The Schur-Zassenhaus Theorem)**.** [57] *Let E be a finite group with a normal abelian subgroup A such that $(\sharp A, \sharp E/A) = 1$. Then E is a semi-direct product of A by E/A and any two complements of A in E are conjugate.*

Proof. Using the Lemma, $H^2(E/A, A) = 0$. Using Theorem 40.4.1, it follows that E is a semi-direct product, as stated. Moreover, the conjugacy classes of the complements are classified by $H^1(E/A, A)$ that, by the same lemma, is trivial too. □

Example 40.4.4. Consider the Corollary for the case of the alternating group A_4 and its Kline subgroup $K = \{1, (12)(34), (13)(24), (14)(23)\}$, which is a normal abelian subgroup. The corollary implies that A_4 is a semi-direct product of the form $K \rtimes H$. This is true, we can take as H the subgroup generated by any 3-cycle. Concretely, we can take $H = \langle(123)\rangle$. Moreover, any two such choices of H are conjugate in A_4 as subgroups. This is also easy to check even though not every two 3-cycles are conjugate in A_4.

Example 40.4.5. Groups of order $11^2 \cdot 3 \cdot 5 = 1815$. Let E be a group of order 1815. By Sylow's theorem, the 11-Sylow subgroup is normal in G. Call it A. Since the order of A is the square of a prime, A is abelian. The quotient group E/A is a group of order 15, which is of the type $pq, p \nmid (q-1)$, thus is cyclic of order 15. Thus, every group of order 1815 is an extension

$$0 \to A \to E \to \mathbb{Z}/15\mathbb{Z} \to 0.$$

Using the Schur-Zassenhaus Theorem, we conclude that E is necessarily a semi-direct product

$$E \cong A \rtimes_\theta \mathbb{Z}/15\mathbb{Z}, \qquad \theta\colon \mathbb{Z}/15\mathbb{Z} \to \mathrm{Aut}(A).$$

[57] The Theorem is true even if A is not abelian, but is harder.

Moreover, any two complements to A in E are conjugate. Depending on the structure of A, one can say more. We leave the details as an exercise.

(1) If A is a cyclic, thus isomorphic to $\mathbb{Z}/121\cdot\mathbb{Z}$, the group $\mathbf{Aut}(A)$ has cardinality $\varphi(121)=110$. This number is not divisible by 3 and that implies that the action of the subgroup of order in 3 in $\mathbb{Z}/15\mathbb{Z}$ on A is trivial. In this case one can conclude that the 3-Sylow subgroup P_3 of E is normal, and taking any 5-Sylow subgroup P_5, we get a subgroup P_3P_5 of order 15 of E that is a complement to A.
(2) If A is not cyclic, thus isomorphic to $(\mathbb{Z}/11\mathbb{Z})^2$, the group $\mathbf{Aut}(A)$ is isomorphic to $\mathbf{GL}_2(\mathbb{Z}/11\mathbb{Z})$ and has order 120×110, a number divisible by 15. In fact, one can prove that $\mathbf{GL}_2(\mathbb{Z}/11\mathbb{Z})$ has an element of order 15 ($\mathbb{F}_{121}^\times$, which has an element of order 15, acts faithfully on $\mathbb{F}_{121}$ which is isomorphic as a $\mathbb{Z}/11\mathbb{Z}$-vector space to $(\mathbb{Z}/11\mathbb{Z})^2$. Thus, $\mathbb{F}_{121}^\times\subset\mathbf{GL}_2(\mathbb{Z}/11\mathbb{Z})$). In this case, it is possible to have semi-direct product relative to $\theta\colon \mathbb{Z}/15\mathbb{Z}\hookrightarrow\mathbf{Aut}(A)$, such that A has a complement (thus E will have a subgroup of order 15, necessarily cyclic), even though neither a 3-Sylow, nor a 5-Sylow is normal.

41. Calculating group cohomology

In this section we present, without proofs, some general techniques to calculate group cohomology. The proofs use further techniques in homological algebra (such as spectral sequence) and substantial new additions to the theory (restriction-inflation morphisms), use of topological methods, and so on. Nonetheless, there is much value in being aware of such results and understanding how to apply them.

41.1. **5 term exact sequence.** Let G be a finite group and $N\triangleleft G$ a normal subgroup. Let A be a G-module. The group G acts on $A^N:=\{a\in A: n*a=a,\forall n\in N\}$, but N acts trivially. Thus, A^N is a G/N module. As well, every element in G acts as an automorphism of N; $g\in G$ takes $n\in N$ to gng^{-1}. This induces a map

$$H^i(N,A)\to H^i(N,A),$$

which on the level of cocycles takes f to ${}^gf$, where

$${}^gf(x_1,\dots,x_i)=gf(g^{-1}x_1g,\dots,g^{-1}x_ig).$$

Thus, all $H^i(N,A)$ are G-modules.

Theorem 41.1.1. *In the notation above:*

(1) There is a 5 **term exact sequence**

$$0\to H^1(G/N,A^N)\to H^1(G,A)\to H^1(N,A)^G\to H^2(G/N,A^N)\to H^2(G,A).$$

(2) Assume that $H^i(N,A)=0$ *for all* $1\le i\le n-1$. *Then, there is an exact sequence*

$$0\to H^n(G/N,A^N)\to H^n(G,A)\to H^n(N,A)^G\to H^{n+1}(G/N,A^N)\to H^{n+1}(G,A).$$

For examples, see §42.5. The proof can be found either in Rotman: *Introduction to homological algebra*, or in Cartan-Eilienberg: *Homological algebra*.

41.2. **Sylow subgroups and group cohomology.** Let G be a finite group and for each prime p dividing the order of G choose some p-Sylow subgroup G_p of G. For any G-module A there is a restriction map $H^n(G,A)\to H^n(G_p,A)$ and it turns out that its image is contained in a particular subgroup denoted $H^n(G_p,A)^G$. We will not describe this subgroup here, but note that typically G_p is not normal in G and so in spite of the notation these are *not* the G invariant elements; they are called the G-**stable elements**.

Theorem 41.2.1. *There is an isomorphism*

$$H^n(G,A) \cong \oplus_{p|\sharp G} H^n(G_p,A)^G, \qquad n \geq 0.$$

In this decomposition $H^n(G_p,A)^G$ corresponds to the p-primary part of $H^n(G,A)$.

For the proof, see Cartan & Eilenberg, *Homological algebra*, Princeton 1999, Theorem 10.1.

41.3. **Topological methods.** Let G be a group. A $K(G,1)$-**space** for G is a connected path-connected topological space X, with some choice of base point $x_0 \in X$ such that the fundamental group $\pi_1(X,x_0) \cong G$ and for all $n > 1$ the higher homotopy groups $\pi_n(X,x_0) = 0$. An example of such a space is the classifying space BG of G. Such a classifying space can often be gotten by finding a contractible topological space $\tilde{X}$ on which G acts freely and discreetly and letting $X = \tilde{X}/G$.

Theorem 41.3.1. *If X is a $K(G,1)$-space, then the singular (Betti) cohomology of X is isomorphic to the group cohomology:*

$$H^n_{sing}(X,\mathbb{Z}) \cong H^n(G,\mathbb{Z}).$$

We stated the theorem for coefficients in $\mathbb{Z}$, viewed as a trivial G-module. The theorem actually extends to $H^n(G,A)$ for any G-module A, but the formulation requires the introduction of the concept of a *local system* on X, which would take us too far a field.

Example 41.3.2. Let F_b be the free group on $b \geq 1$ generators. Let $\tilde{X}$ be the infinite regular tree of degree b. It has a discrete action of F_b, $X := \tilde{X}/F_b$ is a bouquet of b-circles and is a $K(F_b,1)$ space. For dimension reasons, all $H^n(X,\mathbb{Z})$ vanish for $n \geq 1$. And so we find that

$$H^n(F_b,\mathbb{Z}) \cong \begin{cases} \mathbb{Z} & n = 0 \\ \mathbb{Z}^b & n = 1 \\ 0 & n \geq 2. \end{cases}$$

41.4. **Presentations and group cohomology.** Let G be a finite group of order d. Find a presentation of G:

$$G \cong F_b/R.$$

Since R is a finite index subgroup of F_b, R is free itself, in fact of degree $(b-1)d+1$ (Schreier's Theorem). We thus have an exact sequence of groups

$$1 \to R \to F_b \to G \to 1.$$

We can then calculate $H^1(G,\mathbb{Z}) = \mathrm{Hom}(G,\mathbb{Z}) = 0$ and for $H^2(G,\mathbb{Z})$ we have by Theorem 41.1.1

$$\begin{array}{ccccccccc}
0 & \longrightarrow & H^1(G,\mathbb{Z}) & \longrightarrow & H^1(F_b,\mathbb{Z}) & \longrightarrow & H^1(R,\mathbb{Z})^G & \longrightarrow & H^2(G,\mathbb{Z}) \longrightarrow H^2(F_b,\mathbb{Z}) \\
\| & & \| & & \| & & \| & & \| \quad\quad\quad \| \\
0 & \longrightarrow & 0 & \longrightarrow & \mathbb{Z}^b & \longrightarrow & (\mathbb{Z}^{(b-1)d+1})^G & \longrightarrow & H^2(G,\mathbb{Z}) \longrightarrow 0
\end{array}$$

We conclude that

$$H^2(G,\mathbb{Z}) \cong (\mathbb{Z}^{(b-1)d+1})^G/\mathbb{Z}^b.$$

The action of G on $\mathbb{Z}^{(b-1)d+1}$ is derived by the action of F_b on R by conjugation, while remembering that this group is really $H^1(R,\mathbb{Z}) = \mathrm{Hom}(R,\mathbb{Z})$. This is a very effective method for calculating $H^2(G,\mathbb{Z})$ – in

fact, it is implemented in various mathematical softwares, such as GAP – but, the calculation by hand might be out of reach, as typically we need many generators and relations to present a group G.

Example 41.4.1. Take $G = \mathbb{Z}/2 \times \mathbb{Z}/2$. It has the presentation

$$\langle x, y | x^2, y^2, xyx^{-1}y^{-1} \rangle.$$

Recall that R is the minimal *normal* subgroup of F_2 containing $x^2, y^2, [x, y]$, and by Schreier's theorem, R is free of rank 5 (since R has index 4). We leave it as an exercise that R is generated by

$$x^2, yx^2y^{-1}, y^2, xy^2x^{-1}, [x, y].$$

Every homomorphism of R to $\mathbb{Z}$ vanishes on $[x, y]$. If it is fixed by the action of F_2 then it receives the same value on x^2, yx^2y^{-1} and also the same value on y^2, yx^2y^{-1}. Thus, such a homomorphism f is determined uniquely by its values $f(x^2)$ and $f(y^2)$. But, if f is obtained by restriction from $f\colon F_2 \to \mathbb{Z}$ then $f(x^2)$ and $f(y^2)$ are even. We thus find,

$$H^2(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

42. wallpaper groups and cohomology

◇ This whole section is still in preliminary form. Watch out for possible typos (and let me know!). It is best to combine this section with the scan of my handwritten notes that will be uploaded to myCourses soon. ◇

In this section we consider the problem of classification of wallpaper groups. These are groups that arise as symmetry groups of planar designs (such as ceramic walls, or cloth). It is a very classical topic in the study of symmetry groups, often approached in a somewhat ad hoc manner. Here, following Morandi and Iversen, we use systematically representation theory and group cohomology. The final result is that up to isomorphism there are 17 wallpaper groups.

42.1. **Rigid transformations of $\mathbb{R}^2$.** A function $T : \mathbb{R}^2 \to \mathbb{R}^2$ is called a **rigid transformation** if

$$\|Tx - Ty\| = \|x - y\|, \quad \forall x, y \in \mathbb{R}^2.$$

These form a group that we denote $\mathrm{Isom}(\mathbb{R}^2)$. Let $M : \mathbb{R}^2 \to \mathbb{R}^2$ be an orthogonal linear transformation. If we think of M as a $2x2$ matrix then it is orthogonal if and only if ${}^tM \cdot M = I_2$. Let $v \in \mathbb{R}^2$ be any vector and define

$$T\colon \mathbb{R}^2 \to \mathbb{R}^2, \quad T(x) = Mx + v.$$

The function T is a rigid transformation and one can prove that every rigid transformation of $\mathbb{R}^2$ is of this form. Consequently, there is a split exact sequence

$$1 \to \mathbb{R}^2 \xrightarrow{t} \mathrm{Isom}(\mathbb{R}^2) \xrightarrow{\pi} O_2(\mathbb{R}) \to 1,$$

where the map t takes a vector v to t_v, where t_v is the **translation-by-v** map,

$$t_v(x) = x + v.$$

If we represent an element of $\mathrm{Isom}(\mathbb{R}^2)$ as (v, M) then the multiplication law is

$$(v_1, M_1)(v_2, M_2) = (v_1 + M_1v_2, M_1M_2).$$

Given any subgroup Γ of $\mathrm{Isom}(\mathbb{R}^2)$, define its **translation subgroup** Γ_τ and **point group** $\Gamma_\circ$ by

$$\Gamma_\tau := \{t_v \in \Gamma\}, \quad \Gamma_\circ = \pi(\Gamma).$$

Thus, Γ_τ are the translation maps that lie in Γ and $\Gamma_\circ$ is the projection of Γ on the orthogonal group $O_2(\mathbb{R})$. We would usually think of Γ_τ as a subgroup of $\mathbb{R}^2$, by means of the map $v \mapsto t_v$. One has an exact sequence

$$1 \to \Gamma_\tau \to \Gamma \to \Gamma_\circ \to 1.$$

42.2. **Wallpapers and wallpapers groups.** A **wallpaper** is a subset Ω of $\mathbb{R}^2$. The symmetry group of Ω is denoted

$$\Gamma = \Gamma^\Omega = \{T \in \mathrm{Isom}(\mathbb{R}^2) : T(\Omega) = \Omega\}.$$

We say that Γ is a **wallpaper group**[58] if Γ_τ is a discrete rank 2 subgroup of $\mathbb{R}^2$ and $\Gamma_\circ$ is a finite group; with an appropriate topology on $\mathrm{Isom}(\mathbb{R}^2)$ this is the same as saying that Γ is a discrete co-compact subgroup. The group $\Gamma_\circ$ is often called the **point group**. We have an exact sequence

$$1 \to \Gamma_\tau \to \Gamma \to \Gamma_\circ \to 1.$$

Furthermore, as Γ_τ is abelian, $\Gamma_\circ$ acts on Γ_τ, via conjugation in Γ, and this action is induced by restriction from the action of $O_2(\mathbb{R})$ on $\mathbb{R}^2$. In particular, Γ defines a class in

$$H^2(\Gamma_\circ, \Gamma_\tau) \cong H^2(\Gamma_\circ, \mathbb{Z}^2),$$

for this particular action of $\Gamma_\circ$ on Γ_τ (or on $\mathbb{Z}^2$, once we have chosen an isomorphism $\Gamma_\tau \cong \mathbb{Z}^2$). Note that this is a faithful *rational* (even *integral*) representation of $\Gamma_\circ$ on $\Gamma_\tau \cong \mathbb{Z}^2$.

Conversely, suppose that we have an integral faithful action of a finite group $\Gamma_\circ$ on $\mathbb{Z}^2$ and a class $\zeta \in H^2(\Gamma_\circ, \mathbb{Z}^2)$. We claim that it arises this way. First, take an inner product on $\mathbb{Z}^2$ and average it over the group just as in the proof of Maschke's Theorem 10.2.2 to conclude that there is an inner product on $\mathbb{R}^2$ that is preserved by $\Gamma_\circ$. As all inner products on $\mathbb{R}^2$ are related by change of coordinates, we can conjugate $\Gamma_\circ$ into $O_2(\mathbb{R})$, but our lattice $\mathbb{Z}^2$ is conjugated as well and becomes some general lattice, call it Γ_τ, of rank 2 in $\mathbb{R}^2$, but still equipped with an action of $\Gamma_\circ$ that is now induced by restricting the natural action of $O_2(\mathbb{R})$. An exercise on the exercise list proves that we can find a wallpaper Γ with the specified Γ_τ and $\Gamma_\circ$.

42.3. **Examples.** Here are some examples of wallpapers Ω. Imagine the patterns as extending indefinitely. The names attached to them actually provide a description of the isomorphism type of the groups Γ^Ω, but we will not explain the notation here.

Refer to the fleurs-de-lys wallpaper with the designation pm. With the expected choice of coordinates it has the translation group $\mathbb{Z}^2$. It also have the reflection $(x, y) \mapsto (-x, y)$, given by the orthogonal matrix $\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. We also see the reflection around the vertical line with x coordinate $1/2$. This is the transformation $(x, y) \mapsto (1 - x, y) = (-x, y) + (1, 0)$ and so is the composition of the previous reflection and the translation map $t_{(1,0)}$. One finds that the wallpaper group is a (split) extension

$$1 \to \mathbb{Z}^2 \to \Gamma \to \langle \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \rangle \to 1.$$

Consider now the wallpaper designated $p3m1$ and take the middle point of one of the Y as 0. We see translations by the lattice generated by $(1, 0)$ and $(1/2, \sqrt{3}/2) = (\cos(60^\circ), \sin(60^\circ))$. We can also rotate by 120° and reflect through the x-axis. We thus find that the wallpaper group Γ is a split extension

$$1 \to \langle (1, 0),\ (1/2, \sqrt{3}/2) \rangle \to \Gamma \to \Gamma_\circ \to 1,$$

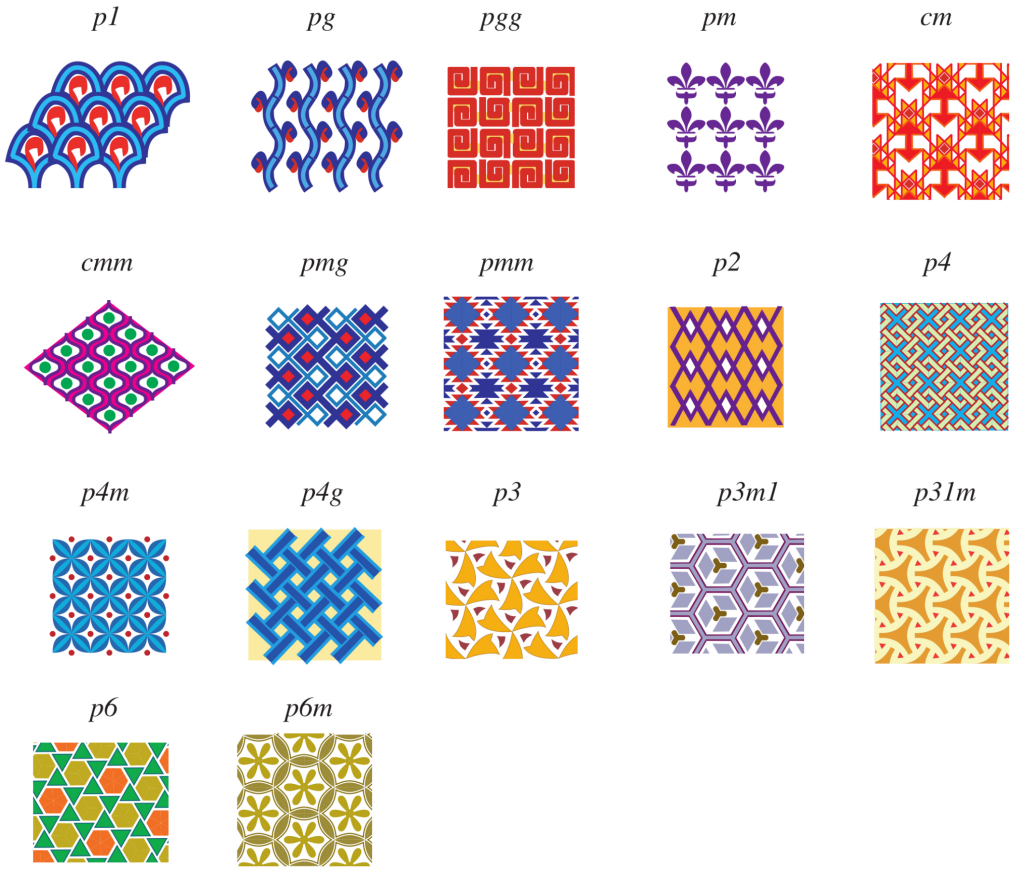
where $\Gamma_\circ \cong D_3$ is the subgroup of the orthogonal group generated by the matrices

$$\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right), \quad \left(\begin{smallmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{smallmatrix}\right).$$

Note that if we take $(1, 0)$ and $(1/2, \sqrt{3}/2)$ as a basis for Γ_τ, then the action of $\Gamma_\circ$ in this basis is

$$\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right), \qquad \left(\begin{smallmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right).$$

[58] Especially in higher dimensions, they are also called "space groups", "crystallographic groups" and "Bieberbach groups", although some authors reserve the latter two names for wallpaper groups with additional properties.



42.4. **The action of $\Gamma_\circ$ on Γ_τ.** As we have already remarked, since Γ is an extension of Γ_τ by $\Gamma_\circ$, $\Gamma_\circ$ acts on Γ_τ by an action induced by conjugation in the extension Γ. This gives a faithful rational representation of $\Gamma_\circ$. Conversely, suppose we are given a faithful rational representation of some finite group $\Gamma_\circ$, $\rho\colon \Gamma_\circ \to \mathrm{GL}_2(\mathbb{Q})$.

Lemma 42.4.1. *The exists a lattice Λ in $\mathbb{R}^2$ preserved by the action of $\Gamma_\circ$.*

Proof. Consider $\Lambda = \sum_{g\in\Gamma_\circ} \rho(g)(\mathbb{Z}^2)$. Note that each $\rho(g)(\mathbb{Z}^2)$ is a lattice commensurable with $\mathbb{Z}^2$. Thus, Λ is a lattice in $\mathbb{R}^2$ and is preserved by the action of $\Gamma_\circ$. □

Thus, by means of Λ, and a basis of Λ, we can view $\Gamma_\circ$ as a subgroup of $\mathrm{GL}_2(\mathbb{Z})$ by $\rho\colon \Gamma_\circ \to \mathrm{GL}_2(\mathbb{Z})$. We have already discussed the following Lemma.

Lemma 42.4.2. *There is an inner product on $\mathbb{R}^2$ preserved by G.*

As there is a change of basis turning this inner product into the standard one, we see that we can conjugate the representation ρ so that we get a representation

$$\rho\colon \Gamma_\circ \to O_2(\mathbb{R}).$$

After this change of coordinates $\Gamma_\circ$ acts on the lattice Λ (whose new coordinates need not even be rational) by a faithful action, compatible with the action of $O_2(\mathbb{R})$.

42.5. **The classification of wallpaper groups.** The main idea is that every wallpaper group Γ is an extension of a lattice $\Lambda = \Gamma_\tau$ by a finite group $G = \Gamma_\circ < O_2(\mathbb{R})$ that satisfies the crystalographic restrictions. As in the Exercise List, we have a list of those. Thus,

- For each finite group G satisfying the crystalogrphic restrictions, namely for a group G among the cyclic groups C_1, C_2, C_3, C_4, C_6 or the dihedral groups D_1, D_2, D_3, D_4, D_6, find its faithful rational representations $G \to \mathrm{GL}_2(\mathbb{Q})$ and for each find the integral lattices preserved by G. This gives 9 possible groups (as $C_2 \cong D_1$) with 13 possibilities of integral representations.
- For every such action $\rho\colon G \to \mathrm{GL}_2(\mathbb{Z})$, calculate $H^2(G, \mathbb{Z}^2)$. Since we can realize the action of G on $\mathbb{Z}^2$ as an action of a subgroup $G_\circ$ of $O_2(\mathbb{R})$ on a lattice in $\mathbb{R}^2$, it follows from the assignments that every such extension arises as a wallpaper group in $\mathbb{R}^2$. This gives 18 extension classes in total.
- Determine which extensions are actually isomorphic as groups. This gives 17 wallpaper groups up to isomorphism. In the case of wallpaper groups one can just do it "by hand", but it is good to be aware of a the following theorem that holds in $\mathbb{R}^n$ too:

Theorem 42.5.1 (Zassenhaus)**.** *Through the faithful action of G on $\mathbb{Z}^2$ view G as a subgroup of $\mathrm{GL}_2(\mathbb{Z})$ and let N be its normalizer in $\mathrm{GL}_2(\mathbb{Z})$. Then N acts on $H^2(G, \mathbb{Z}^2)$ by $\zeta \mapsto {}^n\zeta, n \in N$, where*

$$ {}^n\zeta(x, y) = n\zeta(n^{-1}xn, n^{-1}yn). $$

Two extensions of $\mathbb{Z}^2$ by G, corresponding to classes $\zeta, \eta \in H^2(G, \mathbb{Z}^2)$, are isomorphic as abstract groups if and only if ζ, η are in the same orbit of N.

Group G	$H^2(G, \mathbb{Z}^2)$	$\sharp\, H^2(G, \mathbb{Z}^2)$	wallpaper
C_1	0	1	p1
C_2	0	1	p2
C_3	0	1	p3
C_4	0	1	p4
C_6	0	1	p6
$D_{1,p}$	$\mathbb{Z}/2\mathbb{Z}$	2	pm, pg
$D_{1,c}$	0	1	cm
$D_{2,p}$	$(\mathbb{Z}/2\mathbb{Z})^2$	4	pmm, pmg, pgg
$D_{2,c}$	0	1	cmm
$D_{3,\ell}$	0	1	p3m1
$D_{3,s}$	0	1	p31m
D_4	$\mathbb{Z}/2\mathbb{Z}$	2	p4m, p4g
D_6	0	1	p6m

We do not provide here complete description of how the entries in the table are calculated, but hopefully the reader can either fill in the blanks themselves, or consult the literature. All the actions of the cyclic groups are obtained as rotation actions preserving a suitable lattice Λ depending on the group. One proves in this case that if G is C_3, C_4 or C_6, we take the shortest vector t_1 in the lattice, we let g be a generator of the cyclic group and we let $t_2 = \rho(g)(t_1)$, then t_1, t_2 form a basis for the lattice and the action of g is easily written. For example, suppose that the group is C_3, then we can take the lattice discussed above for the example $p3m1$, and we get that a suitable generator acts by $\left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. In all cases of a cyclic group G the calculation of $H^2(G, \mathbb{Z}^2)$ is quite easy using the results on cohomology of cyclic groups.

The same group $C_2 = \langle g \rangle$ can also act on $\mathbb{Z}^2$ by g acting by $\left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, or by g acting by $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. In the first case, we denote the group $D_{1,p}$ and in the second case we denote it $D_{1,c}$. Both are cyclic groups of

order 2, but the designations allow us to keep track on how they act. We can easily calculate the cohomology group $H^2(G,\mathbb{Z}^2)$ in all cases G is cyclic.

Similarly, $D_{2,p}$ and $D_{2,c}$ are both the dihedral group $D_2 \cong (\mathbb{Z}/2\mathbb{Z})^2$ but with different integral representations. Ditto for $D_{3,p}$ and $D_{3,c}$.

How do we calculate the cohomology for these dihedral groups?? As the groups are not cyclic anymore, this is a non-trivial calculation. One can use the methods of § 41.

Consider for example the case of the group $D_{2,c}$ realized as a subgroup of $O_2(\mathbb{R})$ by

$$\{\pm I_2, \pm \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\}$$

Let us take the cyclic group $N = \{I_2, \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\}$ of $D_{2,c}$. Using the formula for cohomology of cyclic groups, the operator η is

$$\eta = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right),$$

and it kernel is $\mathbb{Z}(1,-1)$. The operator δ is

$$\delta = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right),$$

and its image is again $\mathbb{Z}(1,-1)$. It follows that $H^1(N,\mathbb{Z}^2) = 0$. Note that $(\mathbb{Z}^2)^N = \mathbb{Z}(1,1)$. We may apply Theorem 41.1.1 (b) to get an exact sequence:

$$0 \to H^2(D_{2,c}/N, \mathbb{Z}(1,1)) \to H^2(D_{2,c},\mathbb{Z}^2) \to H^2(N,\mathbb{Z}^2)^{D_{2,c}}.$$

The first term is a quotient of $\mathbb{Z}(1,1)^{D_{2,c}/N} = ((\mathbb{Z}^2)^N)^{D_{2,c}/N} = (\mathbb{Z}^2)^{D_{2,c}} = 0$ and is therefore 0. The third term is $\mathbb{Z}(1,1)/\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)\mathbb{Z}^2 = 0$. Therefore, also $H^2(D_{2,c},\mathbb{Z}^2) = 0$.

The case of $D_{2,p}$ is harder to calculate and also it turns out that 2 of the 4 extension classes define isomorphic groups, reducing the number of planar crystallographic groups to 17 isomorphism classes.

Let us also consider the case of $G_\circ \cong D_3$. In both realizations of D_3 ($D_{3,\ell}$, $D_{3,s}$) an element of order 3 can be taken as $\left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. In the realization $D_{3,\ell}$ one has a reflection $\left(\begin{smallmatrix} 1 & 0 \\ 1 & -1 \end{smallmatrix}\right)$ and in the realization $D_{3,s}$ one has a reflection $\left(\begin{smallmatrix} 1 & -1 \\ 0 & -1 \end{smallmatrix}\right)$. The direct application of Theorem 41.1.1 runs into difficulties: the only normal subgroup of D_3 is A_3 and its first cohomology $H^1(A_3,\mathbb{Z}^2) \cong \mathbb{Z}/3\mathbb{Z}$ and so is not trivial. But, we can reduce the degree of the homology using the following lemma.

Lemma 42.5.2. *Let Γ be a wallpaper group then*

$$H^2(\Gamma_\circ, \Gamma_\tau) \cong H^1(\Gamma_\circ, \mathbb{R}^2/\Gamma_\tau).$$

Proof. Consider the exact sequence of $G_\circ$-modules:

$$0 \to \Gamma_\tau \to \mathbb{R}^2 \to \mathbb{R}^2/\Gamma_\tau \to 0.$$

Note that $H^n(\Gamma_\circ,\mathbb{R}^2) = 0$ for all $n \geq 1$, because multiplication by $\sharp\Gamma_\circ$ kills it, but at the same time multiplication by $\sharp\Gamma_\circ$ is an isomorphism on $\mathbb{R}^2$ and so induces an automorphism of $H^n(\Gamma_\circ,\mathbb{R}^2)$, so it follows that $H^n(\Gamma_\circ,\mathbb{R}^2) = 0$. Therefore, in the long exact sequence in cohomology we get an exact sequence

$$0 = H^1(G_\circ,\mathbb{R}^2) \to H^1(G_\circ,\mathbb{R}^2/\Gamma_\tau) \to H^2(G_\circ,\Gamma_\tau) \to H^2(G_\circ,\mathbb{R}^2) = 0.$$

And so, the claim of the lemma follows. □

To apply Theorem 41.1.1 we will need to know $(\mathbb{R}^2/\mathbb{Z}^2)^{C_3}$.

Lemma 42.5.3. *We have*

$$(\mathbb{R}^2/\mathbb{Z}^2)^{C_3} \cong H^1(C_3,\mathbb{Z}^2) \cong \mathbb{Z}/3\mathbb{Z},$$

and D_3/C_3 acts on it either trivially, or by multiplication by -1.

Proof. Consider the sequence $0 \to \mathbb{Z}^2 \to \mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2 \to 0$ and take invariants to get

$$0 \to (\mathbb{R}^2/\mathbb{Z}^2)^{C_3} \to H^1(C_3, \mathbb{Z}^2) \to H^1(C_3, \mathbb{R}^2) = 0.$$

Moreover, the operator $\eta = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) + \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\delta = I_2 - \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 2 & 1 \\ -1 & 1 \end{smallmatrix}\right)$. This gives, $H^1(C_3, \mathbb{R}^2) \cong \mathbb{Z}^2[\eta]/\delta\mathbb{Z}^2 \cong \mathbb{Z}/3\mathbb{Z}$ and the group of order 2, D_3/C_3 can only act in one of two ways. □

Applying now Theorem 41.1.1 we find

$$0 \to H^1(D_3/C_3, (\mathbb{R}^2/\mathbb{Z}^2)^{C_3}) \to H^1(D_3, \mathbb{R}^2/\mathbb{Z}^2) \to H^1(C_3, \mathbb{R}^2/\mathbb{Z}_2)^{D_3}.$$

Using again the formula for the cohomology of a cyclic group we find that both $H^1(D_3/C_3, (\mathbb{R}^2/\mathbb{Z}^2)^{C_3})$ and $H^1(C_3, \mathbb{R}^2/\mathbb{Z}_2)$ are 0 and we get that

$$H^1(D_3, \mathbb{R}^2/\mathbb{Z}^2) \cong H^2(D_3, \mathbb{Z}^2) = 0.$$

43. Crossed products and the Brauer group

Let K be a field and L/K a finite Galois extension of degree n with Galois group $G = \mathrm{Gal}(L/K)$. The next theorem connects the Brauer group $\mathrm{Br}(K)$ with group cohomology $H^2(G, L^\times)$. We only prove part of theorem as the full proof requires developing much more of the theory of central simple algebras, including some rather deep and useful results, like the Skolem-Noether Theorem. All this can be found in Jacobson/Basic Algebra II.

Theorem 43.0.1. *There is an injective group homomorphism*

$$H^2(G, L^\times) \to \mathrm{Br}(K),$$

whose image is the subgroup of $\mathrm{Br}(K)$ *consisting of equivalence classes of finite-dimensional central simple* k*-algebras* D *that* **split** *over* L*, meaning such that* $D \otimes_K L \cong M_m(L)$ *for some integer* m*.*

Remark 43.0.2. As L varies over all finite Galois extensions, one gets every element of $\mathrm{Br}(K)$ this way. In fact, if D is a fdCS K-algebra and L is a maximal commutative subfield of D then D splits over L: $D \otimes_K L \cong M_m(L)$ for some integer m.

As said, the proof is long and complicated and uses many results that we didn't prove. Here we only show that there is a well-defined map

$$H^2(G, L^\times) \to \mathrm{Br}(K),$$

as we hope this would de-mystify the statement of the Theorem.

Let

$$f \colon G^2 \to L^\times$$

be a 2-cochain. That means that f satisfies the relations

$$^{\alpha}f(\beta,\gamma) \cdot f(\alpha\beta,\gamma)^{-1} \cdot f(\alpha,\beta\gamma) \cdot f(\alpha,\beta)^{-1} = 1, \qquad \alpha,\beta,\gamma \in G. \tag{41}$$

Introduce a formal symbol u_α of every $\alpha \in G$. Consider the vector space

$$B = \oplus_{\alpha \in G} L u_\alpha.$$

It is a vector space of dimension n over L (hence of dimension n^2 over K) with basis $\{u_\alpha : \alpha \in G\}$. To define multiplication on B we define

(1) $u_\alpha u_\beta = f(\alpha,\beta)u_{\alpha\beta}, \quad \alpha,\beta \in G.$
(2) $u_\alpha \ell = \alpha(\ell)u_\alpha, \quad \alpha \in G, \ell \in L.$

To prove that this is operation is associative, the key point is to check that $u_\alpha(u_\beta u_\gamma) = (u_\alpha u_\beta)u_\gamma$. We have $u_\alpha(u_\beta u_\gamma) = u_\alpha f(\beta,\gamma)u_{\beta\gamma} = {}^\alpha f(\beta,\gamma)u_\alpha u_{\beta\gamma} = {}^\alpha f(\beta,\gamma)f(\alpha,\beta\gamma)u_{\alpha\beta\gamma}$. On the other hand, $(u_\alpha u_\beta)u_\gamma = f(\alpha,\beta)u_{\alpha\beta}u_\gamma = f(\alpha,\beta)f(\alpha\beta,\gamma)u_{\alpha\beta\gamma}$. The identity

$$ {}^\alpha f(\beta,\gamma)f(\alpha,\beta\gamma) = f(\alpha,\beta)f(\alpha\beta,\gamma),$$

follows from (41).

We claim that the identity element of this group is

$$e := f(1,1)^{-1}u_1.$$

We note that $f(1,1)^{-1}u_1 \cdot u_\alpha = f(1,1)^{-1}f(1,\alpha)u_\alpha$ and $u_\alpha \cdot f(1,1)^{-1}u_1 = {}^\alpha f(1,1)^{-1}f(\alpha,1)u_\alpha$. So we need to show that

$$f(1,\alpha) = f(1,1), \quad f(\alpha,1) = {}^\alpha f(1,1).$$

This would imply our assertion that e To get the first put $\alpha = \beta = 1$ in (41), giving us $f(1,\gamma) = f(1,1)$ for any γ; the second is obtained by taking $\beta = \gamma = 1$.

We claim that the center of B is precisely Ke. As the Galois group acts trivially on K, multiplication by Ke commutes with every u_α so Ke is contained in $Z(B)$. To show that the center is precisely Ke we prove the following. View the field L as contained in B by $Le = Lu_1, \ell \mapsto \ell e$. Note this is really an isomorphism of fields as $\ell_1 e \ell_2 e = \ell_1 \ell_2 e^2 = \ell_1 \ell_2 e$. It will be enough to prove that L is equal to its centralizer in B. Let $x = \sum_\alpha \ell_\alpha u_\alpha \in B$ and $\ell e \in Le$. Then

$$\ell e x = \sum_\alpha \ell \ell_\alpha u_\alpha, \quad x\ell e = \sum_\alpha \ell_\alpha u_\alpha \ell e = \sum_\alpha \ell_\alpha {}^\alpha \ell u_\alpha.$$

Thus,

$$\ell_\alpha(\ell - {}^\alpha \ell) = 0, \quad \forall \alpha \in G.$$

If $\ell_\alpha \neq 0$ for some $\alpha \neq 1$, we can find an element $\ell \in L$ such that ${}^\alpha \ell \neq \ell$ and get a contradiction. Thus, $x \in L$. As the center of B is contained in the centralizer of L, $Z(B) \subset L$. But an element of $\ell \in L$ that is in the center satisfies $\ell u_\alpha = u_\alpha \ell$ and that implies ${}^\alpha \ell = \ell, \forall \alpha \in G$ and so that $\ell \in K$.

We next claim that B is a simple K algebra. We view now L and K as contained in B, as we have done above. Let I be a two-sided proper ideal of B and consider the algebra $\bar{B} = B/I$. The composition $L \to \bar{B}$ is necessarily an injection (as $\bar{B}$ is not zero and and homomorphism of rings from a field L into a non-zero ring is injective). Let $\bar{u}_\alpha$ denote the image of the element u_α in $\bar{B}$. It is enough to prove that the elements $\{\bar{u}_\alpha : \alpha \in G\}$ are linearly independent over $\bar{L}$. Indeed, this implies that $\dim_{\bar{L}}(\bar{B}) = n$, but on the other hand as $LI \subset I$, I is an L-subspace of B and thus $\dim_L(\bar{B}) = \dim_L(B) - \dim_L(I)$, giving us $\dim_L(I) = 0$ and hence $I = 0$.

The argument is similar to the one used proving *Independence of Characters*. Suppose that the $\{\bar{u}_\alpha\}$ are linearly dependent over $\bar{L}$ and choose a non-trivial linear dependence with the least number of non-zero coefficients:

$$\sum_\alpha \bar{\ell}_\alpha \bar{u}_\alpha = 0.$$

Note that u_α is invertible in B and therefore $\bar{u}_\alpha$ is invertible in $\bar{B}$. By multiplying the equation by a suitable u_α we can arrange that $\ell_1 \neq 0$ and so we assume that henceforth. Note that then there is some $\alpha_0 \neq 1$ such that $\ell_{\alpha_0} \neq 0$ otherwise $\bar{u}_1$ will not be invertible. Choose then some $\ell \in L$ such that ${}^{\alpha_0}\ell \neq \ell$. Then

$$0 = \ell e(\sum_\alpha \bar{\ell}_\alpha \bar{u}_\alpha) - (\sum_\alpha \bar{\ell}_\alpha \bar{u}_\alpha)\ell e = \sum_\alpha \bar{\ell}_\alpha(\ell - {}^\alpha \ell)\bar{u}_\alpha.$$

In this expression if the coefficient of $\bar{u}_\alpha$ was zero, it stays zero, but the coefficient of $\bar{u}_1$ is zero, while the coefficient of $\bar{u}_{\alpha_0}$ is not zero. Namely, we found a non-trivial linear dependence with a smaller number of non-zero coefficients. Contradiction!

We next prove that changing the cocycle f by a coboundary b gives an isomorphic algebra. The coboundary b has the property that

$$b(\alpha,\beta) = {}^{\alpha}h(\beta)h(\alpha\beta)^{-1}h(\alpha),$$

for some function $h\colon G \to L^\times$. Let $v_\alpha = h(\alpha)u_\alpha$. Then

$$v_\alpha v_\beta = h(\alpha)u_\alpha h(\beta)u_\beta = {}^{\alpha}h(\beta)h(\alpha\beta)^{-1}h(\alpha)f(\alpha,\beta)v_{\alpha\beta} = b(\alpha,\beta)f(\alpha,\beta)v_{\alpha\beta}.$$

That shows that in the L-basis $\{v_\alpha\}$ the algebra B has multiplication described by bf. Thus, the isomorphism class of B depends only on the class of f in $H^2(G,L^\times)$.

As said, any other part of the proof is complicated. For example, the statement that the map $H^2(G,L^\times) \to \mathrm{Br}(K)$ is a group homomorphism is not straightforward. On the left hand side we are multiplying cochains to get another cochain and the product still defines an algebra on dimension n^2 over K. On the other hand, on the right hand side we are taking the tensor product over K and the tensor product of two algebras of dimension n^2 over K gives us an algebra of dimension n^4 over K that we need to prove is Brauer-equivalent to the correct algebra of dimension n^2 over K. So there is some work to be done! We refer for all there matters to Jacobson/Basic Algebra II.

However, one consequence of the map being a homomorphism is that the identity element of $H^2(G,L^\times)$, which is represented by the cocycle $f(\alpha,\beta) = 1, \forall \alpha,\beta \in G$, defines an algebra that is Brauer-equivalent to K. Namely, that under our construction, the algebra that we get is simply $M_n(K)$. At least We can check *that*!

Per definition, we have the algebra

$$B = \oplus_{\alpha\in G} L u_\alpha,$$

with the multiplication rules

$$u_\alpha u_\beta = u_{\alpha\beta}, \quad u_\alpha \ell = \alpha(\ell)u_\alpha, \quad \forall \alpha,\beta \in G.$$

We now show that $B \cong M_n(K)$. First, view L as a K vector space of dimension $n = [L:K]$. Then, for every $\ell \in L$ the multiplication-by- ℓ-map

$$[\ell]\colon L \to L, \quad [\ell](x) = \ell x,$$

is a K-linear map and that gives us an inclusion

$$L \hookrightarrow \mathrm{End}_K(L), \quad \ell \mapsto [\ell].$$

In addition, for every $\alpha \in G$ is have the map $L \to L, x \mapsto {}^{\alpha}x$, which is a K-linear map that we denote v_α. Thus,

$$v_\alpha \in \mathrm{End}_K(L).$$

Furthermore, the following relations hold:

$$v_\alpha v_\beta = v_{\alpha\beta}, \quad v_\alpha[\ell] = [{}^{\alpha}\ell]v_\alpha.$$

This implies that we have a well-defined K-algebra homomorphism

$$B \to \mathrm{End}_K(L) \cong M_n(K), \quad \sum \ell_\alpha u_\alpha \mapsto \sum [\ell_\alpha] v_\alpha.$$

Since B is simple, this homomorphism is injective and, by comparing dimensions over K, it is surjective too. It follows that $B \cong M_n(K)$.

Part 6. **Spectral Sequences**

In this Part 6 of the notes, we discuss spectral sequences and some of their applications. Spectral sequences have become an indispensible tool in algebra, geometry and topology.

Spectral sequences are a method to calculate cohomology groups by means of finer and finer approximations. In general, one one gets information about the Jordan-Hölder factors of cohomology groups, but this is often sufficient information for the purposes for which one uses cohomology to begin with. Indeed, one of the most common information one is looking for is whether a cohomology group is zero or not and that can certainly be gleaned from its Jordan-Hölder factors.

Index