Contents

Some acknowledgements and caveats

- Thanks to the students of the course for correcting several errors: Sannidhya Basu, Rishiraj Baul, Subhajit Chakraborty, Shivprateek Das, Amritendu Hait, Ajin Shaji Jose, Aiswarya M, Rajesh Manna, Sahil Reja and Jaskaran Singh.

- I also often followed, at times closely, Arvind Nair's lecture notes, in addition to many standard sources such as Serge Lang's book.
- Not enough proofreading has been done, so there will be numerous inaccuracies. Use at your own risk. The notes for a few of the lectures are incomplete.

1. Lecture 1: Modules over a PID, and applications

(The first 1.5 lectures were probably misguided; I recommend skipping them, and doing this material later in the light of Baer's criterion and the Krull-Schmidt theorem for a more efficient approach.)

1.1. **Background.** For now we will consider only commutative associative rings with 1 unless mentioned otherwise (as we will for occasional digressions).

We will assume basic facts about Euclidean domains, principal ideal domains (PIDs), Unique factorization domains (UFDs) and integral domains. Recall:

$$\text{Euclidean domains } \subset \text{ PIDs } \subset \text{ UFDs } \subset \text{ Integral domains.}$$

These inclusions are all proper: $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$ is a PID that is not a Euclidean domain, $\mathbb{Z}[x]$ or $\mathbb{C}[x,y]$ is a UFD that is not a PID, and $\mathbb{Z}[\sqrt{-5}]$ is an integral domain that is not a UFD. If a, b belong to a PID R, $gcd(a,b)$ will denote a *choice of* a generator for the ideal (a,b).

1.2. **Some theorems associated to PIDs.** The main topic for today's lecture concerns the following theorem:

Theorem 1.1 (Structure theorem for f.g. modules over a PID)**.** *Let M be a f.g. (= finitely generated) module over a PID R. Then there exists a unique (non-strictly) decreasing sequence of proper ideals $(d_1) \supset \cdots \supset (d_n)$ such that:*

$$M \cong \bigoplus_{i=1}^{n} R/(d_i). \tag{1}$$

For an R-module M, where R is any (commutative by convention) ring, let

$$M_{tors} = \{m \in M \mid a \cdot m = 0 \text{ for some non-zero-divisor } a \in R\},$$

a submodule of M, called the torsion submodule of M. Note that in an integral domain, 'non-zero-divisor' is just nonzero. M is said to be a torsion module if $M = M_{tors}$, and torsion-free if $M_{tors} = 0$. It is easy to see that M_{tors} is always a torsion module, and that M/M_{tors} is torsion-free.

Remark 1.2. Before going ahead, some remarks and easy consequences of Theorem 1.1. Assume the setting of the theorem, so in particular M is a f.g. module over the PID R, and we have $d_1|d_2|\ldots|d_n$:

- While the d_i themselves are canonical (as stated in the theorem), the decomposition (1) is highly noncanonical. e.g., $\mathrm{M} = \mathbb{Z} \oplus \mathbb{Z}$ can be identified with $\mathbb{Z} \oplus \mathbb{Z}$ either via the identity map or via the self-isomorphism sending $(1,0)$ to $(1,0)$ and $(0,1)$ to $(1,1)$.

- The theorem gives:

$$M_{tors} \cong \bigoplus_{i:d_i \neq 0} R/(d_i), \qquad M/M_{tors} \cong \bigoplus_{i:d_i=0} R/(d_i) = \bigoplus_{i:d_i=0} R.$$

  In particular, M is torsion-free if and only if $d_i = 0$ for all i, and is torsion if and only if $d_i \neq 0$ for all i.
- The theorem thus implies that any finitely generated torsion-free module over a PID is free.
- The theorem fails for any integral domain that has a finitely generated non-principal ideal I (in other words, when R is not a Bézout domain), such as $\mathbb{Z}[\sqrt{-5}]$: for any such I, and it is easy to see that I is free if and only if it is principal.
- The theorem doesn't generalize in obvious ways to infinitely generated modules over PIDs. For instance, the $\mathbb{Z}$-module $\mathbb{Q}$ is torsion-free, and any two elements in it are linearly dependent, yet it is not isomorphic to $\mathbb{Z}$. A similar comment applies to $\mathbb{Q}/\mathbb{Z}$.

1.3. **Related results.** The crucial input in the proof of the structure theorem for modules over a PID (Theorem 1.1) will be the Smith normal form:

Theorem 1.3. *Let R be a PID, and $A \in M_{m\times n}(R)$. Then $\exists\, S \in GL_m(R)$ and $T \in GL_n(R)$ such that:*

$$S \cdot A \cdot T \qquad \textit{has a form like} \qquad \begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & \ddots & & \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix}, \tag{2}$$

where $d_i | d_{i+1}$ for all $1 \leqslant i \leqslant r$. Here, the above matrix representation shows a situation where $n > m$, and should be analogously modified if $m \leqslant n$: for some $0 \leqslant r \leqslant \min(m, n)$, we have the top left $r \times r$ diagonal block with entries $d_1, \ldots, d_r$, and all remaining entries are zero; by convention, an entry where nothing is written is taken to be 0.

Morever, r and the ideals $(d_i) \subset R$ are uniquely determined.

Remark 1.4. It can be proved without difficulty, but we will skip giving the details, that for each i, $d_i = \delta_i(A)/\delta_{i-1}(A)$, where $\delta_j(A)$ stands for the gcd of all the $j \times j$ minors of A. This property is immediate for a matrix as in (2), so the point is to show that the $\delta_j(A)$ is unchanged when A is right or left multiplied by an invertible matrix. While one can see this using exterior algebra, one can also see this more elementarily, using that right-multiplication (resp., left-multiplication) replaces columns (resp., rows) by their linear combinations, and hence replaces ideals generated by minors with subideals; then use invertibility. For more details, see Lemmas 1 and 2 of the post "Fitting ideals of modules" in Matt Baker's blog: https://mattbaker.blog/2022/11/21/fitting-ideals-of-modules/

Theorem 1.5 (Restatement of the Smith normal form theorem)**.** *Let R be a PID, and let M, N be finitely generated free R-modules. Let $A : N \to M$ be an R-module homomorphism. Then there exist bases $e_1, \ldots, e_n$ of N and $f_1, \ldots, f_m$ of M, and elements $d_1|d_2|\ldots|d_r$ of R for some $0 \leqslant r \leqslant \min(m, n)$, such that for $1 \leqslant i \leqslant n$:*

$$Ae_i = \begin{cases} d_i \cdot f_i, & \text{if } 1 \leqslant i \leqslant r, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Proof. This is a straightforward translation of Theorem 1.3 using (the free module version of) the dictionary between linear transformations and matrices: S and T play the role of change of basis matrices. □

1.4. **The structure theorem from the Smith normal form – I.**

Remark 1.6. To deduce the existence assertion of structure theorem for modules over a PID (Theorem 1.1) from the Smith normal form (Theorem 1.3), it suffices to prove the following two statements:

(i) The existence assertion of Theorem 1.1 holds for finitely presented modules over a PID (see Definition 1.7 below for 'finitely presented').
(ii) Over a PID, every finitely generated module is finitely presented.

Definition 1.7. An R-module M is said to be finitely presented if there exists a surjection $j : R^m \twoheadrightarrow M$ from a finitely generated free module R^m to M, such that $\ker j$ is finitely generated.

Let us rephrase Definition 1.7 to cultivate familiarity with standard mathematical notation. Recall that a chain of maps of R-modules

$$\cdots \to A \xrightarrow{f} B \xrightarrow{g} C \to \ldots$$

is said to be exact at B if $\ker g = \text{image } f$. An exact sequence of R-modules is a chain of maps of R-modules that is exact at each module that is a source of some map in the sequence and a target of another.

Remark 1.8. It follows that an R-module M is finitely presented if and only if there exists an exact sequence

$$R^n \to R^m \to M \to 0.$$

Such a sequence is called a finite presentation of M. Here, note that the exactness at M is equivalent to the surjectivity of $R^m \to M$, and the exactness at R^m implies that the kernel of $R^m \to M$ is generated by n elements.

Another way of saying this is that M is finitely presented if and only if it is isomorphic to the cokernel of a homomorphism of finitely generated free R-modules.

Lemma 1.9. *Assume the existence of the Smith normal form. Then the existence assertion of the structure theorem, Theorem 1.1, is satisifed whenever M is finitely presented.*

Proof. Let $R^n \xrightarrow{A} R^m \to M \to 0$ be a finite presentation of M. Theorem 1.5 allows us to assume without loss of generality that A has the form of the right-hand side of (2). It is then immediate that

$$M \cong \operatorname{coker}(A) \cong R/(d_1) \oplus \cdots \oplus R/(d_r) \oplus R^{m-r} \cong R/(d_1) \oplus \cdots \oplus R/(d_m),$$

where $d_{r+1} = \cdots = d_m = 0$. □

1.5. **The structure theorem from the Smith normal form – II.** To conclude the existence assertion in the structure theorem from the Smith normal form, it suffices to show that finitely generated modules over a PID are also finitely presented: we will see that this is true more generally over Noetherian rings:

Definition 1.10. A (commutative by convention) ring R is said to be Noetherian if any increasing chain of ideals

$$I_1 \subset I_2 \subset \dots$$

stabilizes, i.e., $I_r = I_{r+1} = \dots$ for some large enough r. This is also phrased as saying that R satisfies the ascending chain condition on ideals. More generally, for noncommutative R, we can talk of left-Noetherian rings (resp., right-Noetherian rings) as those that satisfy the ascending chain condition on left-ideals (resp., right-ideals).

Definition 1.11. An R-module M is Noetherian if the collection of its submodules satisfies the ascending chain condition.

Exercise 1.12.
(i) R is a Noetherian ring if and only if every ideal $I \subset R$ is finitely generated. The R-module M is Noetherian if and only if every submodule of M is finitely generated.
(ii) R is Noetherian as a ring if and only if it is Noetherian as a module over itself.
(iii) Any PID is a Noetherian ring.

Lemma 1.13. *Let R be a Noetherian ring. Then every f.g. R-module M is Noetherian (as an R-module).*

Proof. If M is generated by a single element, then $M \cong R/I$ (as an R-module) for some ideal $I \subset R$, and we are done because there is an inclusion-preserving bijection between the submodules of M, and the set of ideals of R containing I: these latter satisfy the ascending chain condition.

The general case is by induction on the number of generators. Let M have n generators $x_1, \dots, x_n$, and assume the lemma to be true for modules with at most $n-1$ generators. Set $M' := Rx_1 \subset M$, and let $M'' = M/M'$. Write $j : M \to M''$ for the obvious map. Then M' and M'' are respectively generated by 1 and $n-1$ elements, and are hence Noetherian.

Now let $M_1 \subset \cdots \subset M_n \subset \dots$ be a chain of submodules of M. Then

$$M_1 \cap M' \subset M_2 \cap M' \subset \dots \qquad \text{and} \qquad j(M_1) \subset j(M_2) \subset \dots$$

are respectively ascending chains of submodules of M' and M''. Since M' and M'' are Noetherian, there exists $r \gg 0$ such that

$$M_r \cap M' = M'_{r+1} \cap M' = \dots \qquad \text{and} \qquad j(M_r) = j(M_{r+1}) = \dots$$

It is enough to show that the inclusion $M_r \subset M_{r+1}$ is an equality: the same argument will imply that $M_{r+1} = M_{r+2} = \dots$. Thus, let $m \in M_{r+1} \backslash M_r$; it suffices to show that $m \in M_r$.

Since $j(m) \in j(M_{r+1}) = j(M_r)$, there exists $n \in M_r$ such that $j(m) = j(n)$. Thus, $m - n \in \ker j = M'$, so that

$$m - n \in M_{r+1} \cap M' = M_r \cap M',$$

so that $m \in n + (M_r \cap M') \subset M_r + (M_r \cap M') = M_r$, as desired. □

Corollary 1.14. *Let R be a Noetherian ring. Then every finitely generated R-module M is finitely presented.*

Proof. If M has m generators, giving us a surjection $A : R^m \twoheadrightarrow M$, then since R^m is Noetherian by Lemma 1.13, $\ker A \subset R^m$ is finitely generated (see Exercise 1.12). Therefore, M is finitely presented. □

Lemma 1.15. *Assume the existence assertion of the Smith normal form (Theorem 1.3). Then the existence assertion of the structure theorem (Theorem 1.1) holds.*

Proof. Given Corollary 1.14 and the observation that any is Noetherian (Exercise 1.12), this follows from Lemma 1.9. □

1.6. **The theorem of elementary divisors.**

Corollary 1.16 (The theorem of elementary divisors)**.** *Let R be a PID. If M is a finitely generated free R-module and $N \subset M$ a submodule, then there exists bases $f_1, \dots, f_m$ for M and $e_1, \dots, e_n$ for N, where $n \leqslant m$, and nonzero elements $d_1 | \dots | d_n$ of R, such that $e_i = d_i \cdot f_i$ for $1 \leqslant i \leqslant n$.*

Proof, assuming Theorem 1.3. By Lemma 1.13, N is finitely generated, in addition to being torsion-free.

Since we are assuming Theorem 1.3, Lemma 1.15 implies that the existence assertion of the structure theorem (Theorem 1.1) holds, so as in Remark 1.2 it follows that N is free as well. This allows us to apply Theorem 1.5 (which is a corollary of Theorem 1.3) to the inclusion morphism $A : M \hookrightarrow N$. We are then done, on observing that the 'r' given by that theorem equals n by the injectivity of A. □

Remark 1.17. In particular, Corollary 1.16 tells us that for a PID R, each submodule of a free R-module of rank m is free of some rank $n \leqslant m$. Thus, for instance, taking $R = \mathbb{Z}$, each subgroup of $\mathbb{Z}^2$ is trivial, or an infinite cyclic group $\mathbb{Z} \cdot a$ for some $a \in \mathbb{Z}^2$, or a free rank two group $\mathbb{Z}a \oplus \mathbb{Z}b \subset \mathbb{Z}^2$ for some linearly independent (nonzero) elements $a, b \in \mathbb{Z}^2$.

1.7. **The existence of Smith normal form.** The main argument we give will follow a proof given in Matt Baker's blog. But before that, we give an idea of a proof in the 2×2 case, this time following the wikipedia.

Slogan: The idea of the proof is that, as in Remark 1.4, 'd_1' should be a gcd of all the entries of A. Because we are working with a PID, we will be able to make manipulations replacing an entry with the gcd of that entry and another. Then somehow induct.

Consider $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(R)$, where R is a PID. We will sucessively multiply this matrix on the left and the right by elements of $GL_2(R)$, so that we finally get a matrix 'in the Smith normal form'. If this matrix is the zero matrix there is nothing to do, so assume it isn't.

Case 1: $a = c = 0$. We have $\sigma, \tau \in R$ such that $b\sigma + d\tau = \beta := gcd(b, d)$. Let $b_1 = b/\beta, d_1 = d/\beta \in R$. Then

$$\begin{pmatrix} \sigma & \tau \\ -d_1 & b_1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & \beta \\ 0 & -d_1 b + b_1 d \end{pmatrix},$$

where we note that $\begin{pmatrix} \sigma & \tau \\ -d_1 & b_1 \end{pmatrix}$ has determinant $\sigma b_1 + \tau d_1 = 1$ and hence belongs to $SL_2(R) \subset GL_2(R)$. By adding $(d_1 b - b_1 d)/\beta$ times the first row to the second – an operation that is implemented by left-multiplication by

$$\begin{pmatrix} 1 & \\ (d_1 b - b_1 d)/\beta & 1 \end{pmatrix} \in M_2(R)$$

– we get $\begin{pmatrix} 0 & \beta \\ 0 & 0 \end{pmatrix}$. Interchanging the columns – which is right-multiplication by $\begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$ – we get to the desired Smith normal form $\begin{pmatrix} \beta & 0 \\ 0 & 0 \end{pmatrix}$. This finishes Case 1, where $a = c = 0$.

Case 2. $a \neq 0$ or $c \neq 0$. For this general case, we will implicitly use the following ideas illustrated in Case 1: the 'row operations' of interchanging rows and adding a multiple of a row to another can be implemented by left-multiplication by an element of $GL_2(R)$, and analogous column operations by right-multiplication by elements of $GL_2(R)$.

Thus, we exchange rows if necessary to assume that $a \neq 0$. First, let us see how to perform operations to ensure $a|b$. If not, let $\beta = gcd(a, c)$, $a\sigma + c\tau = \beta$ for some $\sigma, \tau \in R$, $a_1 = a/\beta \in R$ and $c_1 = c/\beta \in R$. Then we have

$$\begin{pmatrix} \sigma & \tau \\ -c_1 & a_1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \beta & * \\ -c_1 a + a_1 c & *, \end{pmatrix}$$

where as before the left-most matrix lies in $SL_2(R) \subset GL_2(R)$, and '$*$' refers to an entry from R whose precise value we are unconcerned about.

Thus, we can add the product of the first row with $(c_1 a - a_1 c)/\beta \in R$ to the second (a $GL_2(R)$-left-multiplication), to get to the form $\begin{pmatrix} \beta & * \\ 0 & * \end{pmatrix}$. A similar argument involving

right-multiplication reduces us to a diagonal matrix. Now let us re-use the letters a and d, to write the resulting matrix as

$$\begin{pmatrix} a & \\ & d \end{pmatrix}.$$

The problem is that a may not divide d, so the above method may not be in smith normal form. However, we may add the second column to the first, to get $\begin{pmatrix} a & \\ d & d \end{pmatrix}$ and repeat the argument to 'eliminate the lower left entry' seen above, to replace a with $gcd(a, d)$ and then the now-possibly-nonzero lower left and top right entries back to 0. We again get a diagonal matrix. This messes with the value of d, which however can be obtained up to associates because the determinant is preserved up to an element of $R^\times$ (because we are only multiplying by elements of $GL_2(R)$) [1]: a is replaced by (a, d), so d has to be replaced by $ad/(a, d)$, which clearly is a multiple of (a, d). Thus, now the matrix is in the desired Smith normal form. (This step is more complicated for larger matrices since we have lesser control on the other diagonal entries, but one inducts on the number of prime factors of the top-left entry).

Now we give a proof of existence in the general case, shifting to following Matt Baker's blog.

Proof of existence of Smith normal form, general case. For $r \in R$, let $l(r)$ equal the number of prime factors of r, with $l(0)$ being ∞.

Say that $A, A' \in M_{m\times n}(R)$ are equivalent if $A = BA'C$ for some $B \in GL_m(R), C \in GL_n(R)$. We wish to choose an element of the equivalence class of a given matrix $A \in M_{m\times n}(R)$ so as to be as in the right-hand side of (2).

Without loss of generality, we assume $A = [a_{ij}]$ to be chosen so that $l(a_{11}) \leqslant l(a'_{11})$ whenever $A' = [a'_{ij}]$ is equivalent to A. Since, for all $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$ A' can be replaced by an equivalent $A'' = [a''_{ij}]$ with $a''_{11} = a'_{ij}$ (by permuting the i-th and the 1-st rows, and the j-th and the 1-st columns), we conclude that $l(a_{11}) \leqslant l(a'_{ij})$ for all i, j; in particular, $l(a_{11}) \leqslant l(a_{ij})$ for all i, j.

Idea. One hopes that this will force a_{11} to be the gcd of the matrix entries.

Claim. $a_{11}|a_{1j}$ for all j.
Suppose this is not true for some j. To get a contradiction, we may replace A by the matrix obtained by swapping the 2-nd and the j-th columns of A, which satisifes the same condition on a_{11} and is equivalent to A; thus, we may assume that $a_{11} \nmid a_{12}$. We let $\beta = gcd(a_{11}, a_{12})$, choose $\sigma, \tau \in R$ such that $a_{11}\sigma + a_{12}\tau = \beta$, and let $\bar{a}_{11} = a_{11}/\beta, \bar{a}_{12} =$

[1] In fact, up to ± 1: we have only multiplied by elements in either $SL_2(R)$ or have determinant -1, such as antidiag$(1, 1)$

$a_{12}/\beta \in R$. Then consider the following matrix, which is equivalent to A:

$$A' := A \cdot \begin{pmatrix} \sigma & -\bar{a}_{12} & \\ \tau & \bar{a}_{11} & \\ & & I_{n-2} \end{pmatrix} \in M_n(R).$$

Then A' is equivalent to A (the right-most matrix above has determinant 1 and hence belongs to $SL_n(R) \subset GL_n(R)$), but has top-left entry $a_{11}\sigma + a_{12}\tau = \beta$. Since $a_{11} \nmid a_{12}$, $l(\beta) < l(a_{11})$, and since A' is equivalent to A, this contradicts the choice of a_{11}. This proves the claim that $a_{11}|a_{1j}$ for all j.

Similarly, $a_{11}|a_{i1}$ for all $1 \leqslant i \leqslant m$. Now we may do row and column operations involving subtracting a multiple of the first row or column from another row or column, and assume that A takes the form

$$A = \begin{pmatrix} a_{11} & 0_{1\times n-1} \\ 0_{m-1\times 1} & B, \end{pmatrix},$$

where B is an $(m-1)\times(n-1)$-matrix Here, if $m = 1$ or $n = 1$, the form of the matrix will be slightly different, in that it will have only a_{11} and $0_{1\times n-1}$ or a_{11} and $0_{m-1\times 1}$, in which cases we are already in the Smith normal form. Set $d_1 := a_{11}$. By the induction hypothesis (if m or n is greater than 1), we may replace B by an element of $GL_{m-1}(R)BGL_{n-1}(R)$, to assume that A is in the Smith normal form, with a top-left diagonal $(r-1)\times(r-1)$ block $\mathrm{diag}(d_2, \ldots, d_r)$ for some r, and all other entries 0.

Now it suffices to show that $d_1|d_2$. To see this, note A is similar to a matrix of the following form (for simplicity, I am showing only the case where $n > m$):

$$\begin{pmatrix} d_1 & d_2 & & & & \\ & d_2 & & & & \\ & & \ddots & & & \\ & & & d_r & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

But this matrix is similar to A and satisfies the same property of minimizing $l(a_{11}) = l(d_1)$: thus, the 'Claim' in the proof above shows that $d_1|d_2$, as desired. Thus, A is now in the Smith normal form $\square$

Remark 1.18. The approach above, taken from Matt Baker's blog, is more efficient than the one given in the wikipedia page on the Smith normal form, but it does not seem 'constructive' (since one starts with minimizing $l(a_{11})$). At least with the wikipedia approach, if one works with a Euclidean domain, one can work with just elementary row and column operations without involving matrices like $\begin{pmatrix} \sigma & \tau \\ -d_1 & b_1 \end{pmatrix}$.

2. Lecture 2 — More on the structure theorem, applications, and categories and functors

2.1. **Uniqueness in the structure theorem.** Now we will discuss a proof of the uniqueness in the structure theorem for modules over a PID [**Warning: this is very inefficient**]. In other words, R is a PID, we have

$$M \cong \bigoplus_{i=1}^{n} R/(d_i) = \bigoplus_{i=1}^{n'} R/(d_i'),$$

with $(d_1) \supset (d_2) \supset \dots$ and $(d_1') \supset (d_2') \supset \dots$.

We need to show that $n = n'$, and that $(d_i) = (d_i')$ for all i.

Reduction to the case where $M = M_{tor}$.

Note that

$$R^{\#\{i|d_i=0\}} = \bigoplus_{\substack{i=1 \\ d_i=0}}^{n} R/(d_i) = M/M_{tors} = \bigoplus_{\substack{i=1 \\ d_i'=0}}^{n'} R/(d_i') = R^{\#\{i|d_i'=0\}}.$$

Let us show that the above equality implies $\#\{i \mid d_i = 0\} = \#\{i \mid d_i' = 0\}$. Since R is an integral domain, which by definition requires $1 \neq 0$, this follows from the lemma below:

Lemma 2.1. *For this lemma allow R to be an arbitrary commutative ring that is not the zero ring. If $R^{n_1} \cong R^{n_2}$ as R-modules for some nonnegative integers n_1, n_2, then $n_1 = n_2$.*

Proof. Since R is not the zero ring, it has a maximal ideal $\mathfrak{m}$, and the quotient $k := R/\mathfrak{m}$ is a field. Since $R^{n_1} \cong R^{n_2}$ we have

$$\frac{R^{n_1}}{\mathfrak{m}R^{n_1}} \cong \frac{R^{n_2}}{\mathfrak{m}R^{n_2}}.$$

Note that

$$R^{n_i}/\mathfrak{m}R^{n_i} = (R \oplus \cdots \oplus R)/(\mathfrak{m}R \oplus \cdots \oplus \mathfrak{m}R) = (R/\mathfrak{m}R) \oplus \cdots \oplus (R/\mathfrak{m}R) = k^{n_i},$$

so we have $k^{n_1} \cong k^{n_2}$ as R-modules, and hence as vector spaces over $R/\mathfrak{m} = k$. Comparing dimensions, we get $n_1 = n_2$. □

Since this proves that $\#\{i \mid d_i = 0\} = \#\{i \mid d_i' = 0\}$, it is now enough to match up the nonzero (d_i) with the nonzero (d_i'). For this, noting that

$$M_{tors} = \bigoplus_{\substack{i=1 \\ d_i \neq 0}}^{n} R/(d_i) = \bigoplus_{\substack{i=1 \\ d_i' \neq 0}}^{n'} R/(d_i'),$$

we may now replace M with M_{tors}, to assume that

$$M = M_{tors} = \bigoplus_{i=1}^{r} R/(d_i) = \bigoplus_{i=1}^{r'} R/(d_i'),$$

where we now write r and r' for n and n', just to be consistent with the previous lecture's notation. We have $d_1|\dots|d_r$ and $d'_1|\dots|d'_{r'}$.

Claim. $(d_r) = (d'_{r'})$.
Proof. Note that the annihilator of M, namely $\mathrm{Ann}_R(M) := \{a \in R \mid aM = 0\}$, equals (d_r) as well as $(d'_{r'})$. This proves the claim.

Case 1. $(d_r) = (d'_{r'}) = (p)^k$ for a prime $p \in R$ (and $k > 0$, since $M = M_{tors}$).

In this case, for each i, since $d_i|d_r$, we can write $(d_i) = (p^{k_i})$, and similarly $d'_i = (p^{k'_i})$ (so $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_r = k$ and $k'_1 \leqslant k'_2 \leqslant \cdots \leqslant k'_{r'} = k$). Thus, we are given

$$M \cong \bigoplus_{i=1}^{r} R/(p^{k_i}) \cong \bigoplus_{i=1}^{r'} R/(p^{k'_i}). \tag{3}$$

We need to show that $r = r'$ and that $k_i = k'_i$ for each i.

For $a \in R$, let $M[a] = \{m \in M \mid am = 0\}$, an R-submodule of M. For nonnegative integers c, we have $M[p^c] \subset M[p^{c+1}]$, and we will extract information from (3) through its consequence of the form:

$$\frac{M[p^{c+1}]}{M[p^c]} = \bigoplus_{i=1}^{r} \frac{R/(p^{k_i})[p^{c+1}]}{R/(p^{k_i})[p^c]} = \bigoplus_{i=1}^{r'} \frac{R/(p^{k'_i})[p^{c+1}]}{R/(p^{k'_i})[p^c]}. \tag{4}$$

Note that for any nonnegative integer c, $(R/(p^{k_i}))[p^c]$ equals the R-module $(p^{k_i-c})/(p^{k_i})$ if $k_i \geqslant c$, and $R/(p^{k_i})$ if $k_i \leqslant c$. Thus, $(R/(p^{k_i}))[p^{c+1}]/(R/p^{k_i})[p^c]$ is the 0-module if $k_i \leqslant c$, and is isomorphic to $(p^{k_i-c-1})/(p^{k_i-c})$, which is a one-dimensional vector space over $R/(p)$ generated by the image of p^{k_i-c-1}, otherwise (i.e., if $k_i > c$). Thus, the second and third expressions in (6) are vector spaces over $R/(p)$, and equating their dimensions gives:

$$\#\{i \mid k_i > c\} = \#\{i \mid k'_i > c\}.$$

Since this is true for all nonnegative c, we get for all $c \geqslant 1$:

$$\#\{i \mid k_i = c\} = \#\{i \mid k'_i = c\},$$

so that, the d_i being precisely the p^{k_i} and similarly with the d'_i, we get $r = r'$ and $d_i = d'_i$ for all i.

Case 2. $d_r = d'_{r'}$ is general, so up to taking associates, $d_r = d'_{r'} = \prod_{j=1}^{t} p_j^{k_{rj}} = \prod_{j=1}^{t} p_j^{k'_{r'j}}$. Thus, for $1 \leqslant i \leqslant r$, since $d_i|d_r$, we can write $d_i = \prod_{j=1}^{t} p_j^{k_{ij}}$ (up to taking associates). Similarly, for $1 \leqslant i \leqslant r'$, we write $d'_i = \prod_{j=1}^{t} p_{j'}^{k'_{ij}}$. But this time, we need to allow some of the k_{ij} and the k'_{ij} to be 0.

Thus, by the Chinese remainder theorem (reviewed in Remark 2.2 below), we have

$$\bigoplus_{i=1}^{r}\bigoplus_{j=1}^{t} R/(p_j^{k_{ij}}) = \bigoplus_{i=1}^{r} R/(d_i) = M = \bigoplus_{i=1}^{r'} R/(d'_i) = \bigoplus_{i=1}^{r'}\bigoplus_{j=1}^{t} R/(p_j^{k'_{ij}}). \tag{5}$$

It turns out that we can separate the contributions from the different primes p_j: for this, if p and q are different primes in R, we claim that for any $l \geqslant 0$, multiplication by p is an automorphism of $R/(q^l)$. Indeed, this is because $\alpha p + \beta q^l = 1$ for some $\alpha, \beta \in R$, so that multiplication by α serves as an inverse to multiplication by p in $R/(q^l)$. Therefore, for $j \neq l$, $R/(p_l^{k_{il}})$ does not have any nonzero p_j-power torsion, and similarly with the $R/(p_l^{k'_{il}})$, so the argument from Case 1 gives, for each fixed $1 \leqslant j \leqslant t$:

$$\frac{M[p_j^{c+1}]}{M[p_j^c]} = \bigoplus_{i=1}^{r} \frac{R/(p_j^{k_{ij}})[p_j^{c+1}]}{R/(p_j^{k_{ij}})[p_j^c]} = \bigoplus_{i=1}^{r'} \frac{R/(p_j^{k'_{ij}})[p_j^{c+1}]}{R/(p_j^{k'_{ij}})[p_j^c]}. \tag{6}$$

The argument from Case 1 then gives that for each j, the set of *nonzero* k_{ij}'s is the same as the set of nonzero k'_{ij}'s. A little bit more of care is still needed, since some k_{ij}'s and k'_{ij}'s may be zero. But since d_1 is not a unit and is hence divisible by p_j for *some* j, r is the largest possible number of the nonzero k_{ij}'s as j varies from 1 to t. A similar assertion applies to r'. Thus, we conclude that $r = r'$. From this, it follows that for each j, $k_{ij} = k'_{ij}$ for $1 \leqslant i \leqslant r$: this is because the multisets $\{k_{ij} \mid 1 \leqslant i \leqslant r\}$ and $\{k'_{ij} \mid 1 \leqslant i \leqslant r'\}$ have the same number of elements $r = r'$, and the same collections of nonzero elements as observed above, and they both (non-strictly) increase with i. This implies that

$$d_i = \prod_{j=1}^{t} p_j^{k_{ij}} = \prod_{j=1}^{t} p_j^{k'_{ij}} = d'_i.$$

(recall that we had changed them up to associates to get particular prime factorizations; if we hadn't, we could only say $(d_i) = (d'_i)$). This finishes the proof of the uniqueness assertion.

Remark 2.2. We review the Chinese reminder theorem used in the above proof : if $I_1, \ldots, I_r \subset R$ are ideals that are pairwise comaximal, i.e., $I_i + I_j = R$ whenever $i \neq j$, the map $R \to \prod_{i=1}^r R/I_i$ is surjective and quotients to an isomorphism of rings,

$$R/(I_1 \cap \cdots \cap I_r) \to \prod_{i=1}^{r} R/I_i.$$

We have a similar result involving R-modules, except that in the case of modules it is more natural to write $\bigoplus_{i=1}^r$ in place of $\prod_{i=1}^r$.

Remark 2.3. Now we make comments on the above proof of uniqueness.

(i) *Primary decomposition.* For $a \in R$, write $M[a]$ for $\{m \in M \mid am = 0\}$, and $M[a^\infty]$ for $\{m \in M \mid a^j M = 0 \text{ for some } j\}$; these are submodules of M. As above, we write $(d_r) = \mathrm{Ann}_R(M)$, and let $p_1, \ldots, p_t$ be the primes that divide d_r.

As an easy exercise (e.g., follow the 'gcd' argument above) see that we have

$$M = \bigoplus_{j=1}^{t} M[p_j^\infty] = \bigoplus_{p \text{ prime in } R} M[p^\infty] \tag{7}$$

(namely, $M[p^\infty]$ is zero unless $p = p_j$ for some j).

Moreover, the argument of Case 2 gives:

$$M[p_j^\infty] = \bigoplus_{i=1}^{r} R/(p_j^{k_{ij}}).$$

This was essentially what allowed us to reduce to 'Case 1', by separating out the primes $p_1, \ldots, p_t$: namely, each $M[p_j^\infty]$ is canonically determined, the decomposition of M in terms of the $R/(d_i)$ and the $R/(d_i')$ gave us analogous decompositions of $M[p_j^\infty]$ involving the largest powers $p_j^{k_{ij}}$ and $p_j^{k'_{ij}}$ of p_j that divide the d_i and the d_i', and these could be matched up using Case 1.

There is a generalization of (7) to finitely generated modules over arbitrary Noetherian rings, called the *primary decomposition*. However, the precise formulation of this general decomposition is a bit subtler, and not as nice as in (7): getting such a nice decomposition as in (7) does need R to be a PID. You might learn this in your second semester algebra course.

(ii) *Indecomposable modules and the Krull-Schmidt theorem.* Here is what is probably a 'philosophical reason' why we had to work with the $R/(p_j^{k_{ij}})$, and not directly with the $R/(d_i)$. An indecomposable module over a ring R is a module that cannot be written as a direct sum of two proper submodules. [2] According to the Krull-Schmidt theorem, any "finite length" R-module is a direct sum of indecomposable modules, and the isomorphism classes of these indecomposable constituent modules are uniquely determined up to a permutation. Over a PID R, it is an easy exercise to show that the only finitely generated indecomposable R-modules are R and the $R/(p^k)$, p a nonzero prime and k a positive integer. Thus, the Krull-schmidt theorem can be used to replace part of the proof of the uniqueness assertion in the structure theorem. On the other hand, the $R/(d_i)$ are not indecomposable, so there does not seem to be a naive direct way to extract a uniqueness assertion without factorizing the (d_i).

(iii) *Fitting ideals.* Nevertheless, there is another, less naive, approach to show the uniqueness assertion, which does not involve factoring into primes. This involves the notion of Fitting ideals, studied by Hans Fitting. We will give a crude outline here following a blog post of Matt Baker. For each $k \geqslant 0$, define the k-th fitting ideal $Fit_k(M)$ as follows: *choose* a presentation $R^n \xrightarrow{A} R^m \to 0$ of M, and let $Fit_k(M)$ be the ideal generated by the $(m-k) \times (m-k)$-minors of A. Fitting's lemma says that this is independent of the presentation $R^n \xrightarrow{A} R^m \to 0$, i.e., is *intrinsic* to M, justifying the notation $Fit_k(M)$ (without needing to keep track of the presentation in the notation). Since the minors of A determine the d_i, one can use Fitting's lemma to give a proof of the uniqueness assertion in the structure

[2] Note that a simple (or "irreducible") module, namely one that does not have a proper nonzero submodule, is indecomposable, but an indecomposable module may not be simple.

theorem, without using primary decomposition. For more details including a proof of Fitting's lemma, see Matt Baker's post on fitting ideals.

(iv) *Another proof of existence.* For another proof of the existence assertion in the structure theorem, see Serge Lang's 'Algebra'.

(v) *Only the (d_i) are canonical.* 'The copies' of the $R/(d_i)$ in M are not canonical. For instance, when $R = \mathbb{Z}$, $M := \mathbb{Z}/(2) \oplus \mathbb{Z}/(4)$ can be decomposed into $\mathbb{Z}/(2) \oplus \mathbb{Z}/(4)$ in two ways: first in the obvious way, as well as in a second, less obvious, way, in which one sends $(\bar{1}, 0)$ to $(\bar{1}, 0)$ and $(0, \bar{1})$ to $(\bar{1}, \bar{1})$. In the former decomposition of M, the copy of $\mathbb{Z}/(4)$ is generated by $(0, \bar{1})$, and in the latter by $(\bar{1}, \bar{1})$, showing that it is not canonically determined.

2.2. **Application to linear algebra.** Let V be a vector space over a field k, and let $T \in \mathrm{End}_k(V)$, where $\mathrm{End}_k(V)$ is the ring of k-linear endomorphisms of V. We will discuss some 'canonical forms' given by the structure theorem for modules over a PID, which are nice matrix forms of T. We will change notation a bit: now, we will write f_i for what we wrote a d_i before.

Since $T : V \to V$ commutes with scalar multiplication (i.e., $T \in \mathrm{End}_k(V)$), V can be thought of as a $k[x]$-module where x acts by T: concretely, to define a $k[x]$-module structure on T, it is enough to define a homomorphism $k[x] \to \mathrm{End}_{\mathbb{Z}}(V)$, which we can define to be given by the scalar multiplication on k, and so as to send x to T (this is okay, since x is a free variable and T commutes with scalar multiplication). This realizes V as a $k[x]$-module. Since V is finite dimensional while $k[x]$ is infinite dimensional over k, it is immediate that V is torsion as a $k[x]$-module.

Thus, if $f_1 | f_2 | \dots | f_r$ are associated by the structure theorem for modules over a PID to the $k[x]$-module V, we get an isomorphism of $k[x]$-modules:

$$V \cong \bigoplus_{i=1}^{r} k[x]/(f_i). \tag{8}$$

Moreover, this time the f_i can be specified uniquely, by requiring them to be monic.

Remark 2.4. Concretely, (8) means: there exists a k-vector space isomorphism from V to $\bigoplus_{i=1}^{r} k[x]/(f_i)$, such that the transport of $T : V \to V$ under this isomorphism is the k-vector space map $\bigoplus_{i=1}^{r} k[x]/(f_i) \to \bigoplus_{i=1}^{r} k[x]/(f_i)$ given by multiplication by x. Thus, to compute the matrix of T with respect to some basis, it suffices to compute the matrix of multiplication by x with respect to the corresponding basis of $\bigoplus_{i=1}^{r} k[x]/(f_i)$.

Example 2.5. (i) Let $0 \neq f(x) = x^n + a_1 x^{n-1} + \dots + a_n \in k[x]$, and consider the k-vector space $k[x]/(f)$. A basis of this vector space is given by $1, x, \dots, x^{n-1}$, and with respect to this basis, the matrix of the linear map $k[x]/(f) \to k[x]/(f)$ given

by multiplication by x is:

$$\begin{pmatrix} 0 & & & & -a_n \\ 1 & 0 & & & -a_{n-1} \\ & \ddots & \ddots & \vdots & \vdots \\ & & & 0 & -a_2 \\ & & & 1 & -a_1 \end{pmatrix}. \tag{9}$$

(ii) Let $k[x]/(f)$ be as above, but assume that $f(x) = (x-\alpha)^n$. This time, a basis for $k[x]/(f(x))$ can be given as $1, x-\alpha, \ldots, (x-\alpha)^{n-1}$. With respect to this basis, multiplication by x has matrix

$$\begin{pmatrix} \alpha & & & & \\ 1 & \alpha & & & \\ & \ddots & \ddots & & \\ & & & \alpha & \\ & & & 1 & \alpha \end{pmatrix}. \tag{10}$$

Corollary 2.6 (Rational canonical form)**.** *Let k be a field and n a natural number. Then any matrix in $M_n(k)$ can be $GL_n(k)$-conjguated to a matrix in block diagonal form, where each block is as in* (9)*.*

Proof. Follows from Remark 2.4 and the discussion preceding it, together with Example 2.5((i)). □

Corollary 2.7 (Jordan canonical form)**.** *Let k be an algebraically closed field, and n a natural number. Then any matrix in $M_n(k)$ can be $GL_n(k)$-conjguated to a matrix in block diagonal form, where each block is as in* (10)*.*

Proof. Follows from Remark 2.4 and the discussion preceding it, together with the fact that each f_i can be factored as a product $\prod(x-\alpha_{ij})^{n_{ij}}$ since k is algebraically closed, the Chinese remainder theorem to separate the α_{ij} for a given i, and Example 2.5((ii)). □

Remark 2.8. Consider an isomorphism $V \to \bigoplus_{i=1}^{r} k[x]/(f_i)$ as in Remark 2.4, transporting $T \in \mathrm{End}_k(V)$ to multiplication by x. It is immediate then that the minimal polynomial of T is f_r. Further, it is easy to see that the characteristic polynomial of multplication by x on $k[x]/(f_i)$ is f_i (expand the relevant determinant from the top and proceed inductively), so the characteristic polynomial of T is $f_1 f_2 \ldots f_r$.

2.3. **The Cayley-Hamilton theorem.**

Theorem 2.9 (Cayley-Hamilton)**.** *Let R be a (commutative by convention) ring, and $A \in M_n(R)$. If $p(x) = det(xI_n - A)$ is the characteristic polynomial of A, then $p(A) \in M_n(R)$ equals* 0*.*

Proof. We can view R^n as a module over $R[x]$, with x acting via A.

Given an S-module M, there is an obvious structure of an $M_n(S)$-module on $M^n := \bigoplus_{i=1}^n M$, written for convenience as column vectors, according to the usual matrix multiplication: $[a_{ij}] \cdot {}^t(m_1, \ldots, m_n) = (\sum a_{1j}m_j, \ldots, \sum a_{nj}m_j)$. Thus, we get the structure of an $S := M_n(R[x])$-module on $M^n = (R^n)^n$, where $M := R^n$.

It is now easy to check that for this action:

$$\begin{pmatrix} x & & \\ & \ddots & \\ & & x \end{pmatrix} \cdot \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} x \cdot e_1 \\ \vdots \\ x \cdot e_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n a_{i1}e_i \\ \cdots \\ \sum_{i=1}^n a_{in}e_i \end{pmatrix} = {}^tA \cdot \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix},$$

or equivalently:

$$(xI_n - {}^tA) \cdot {}^t(e_1 \ldots e_n) = {}^t(0 \ldots 0).$$

Multiplying by the adjugate matrix of $xI_n - {}^tA$, and using that A and ${}^tA$ have the same characteristic polynomials $p_A = p_{{}^tA}$, we get

$$\begin{pmatrix} p_A(x) & & \\ & \ddots & \\ & & p_A(x) \end{pmatrix} \cdot \begin{pmatrix} e_1 \\ \ddots \\ e_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Thus, $p_A(x) \cdot e_i = 0$ for each i, i.e., $p_A(A) \cdot e_i = 0$ for each i, so that $p_A(A)$ annihilates each e_i, and hence the whole of R^n. In other words, $p_A(A) = 0$. □

Here are steps for an alternate proof, from Arvind's notes:

(i) The assertion is immediate when T is a diagonal matrix, and hence when T is a diagonalizable matrix.
(ii) When $k = \mathbb{C}$, diagonalizable matrices are dense in the space $M_n(k)$ of all matrices, say in the usual topology on $M_n(\mathbb{C}) \cong \mathbb{C}^{n^2}$. Thus, the result holds when $k = \mathbb{C}$.
(iii) $\mathbb{Z}[x_{ij} | 1 \leqslant i, j \leqslant n]$ embeds into $\mathbb{C}$ (use that $\mathbb{C}$ has infinite transcendence degree over $\mathbb{Q}$), using which it is easy to see that the result holds for $\mathbb{Z}[x_{ij} | 1 \leqslant i, j \leqslant n]$.
(iv) Any element of $M_n(R)$ is the image of an element of $M_n(\mathbb{Z}[x_{ij}] | 1 \leqslant i, j \leqslant n)$, so we are done.

2.4. **Categories.**

Not quite a definition 2.10. A category $\mathcal{C}$ consists of:

- A 'collection' $\mathrm{Ob}\,\mathcal{C}$ whose members are called the objects of $\mathcal{C}$; and
- For each X, Y in $\mathrm{Ob}\,\mathcal{C}$, a collection $\mathrm{Mor}_{\mathcal{C}}(X, Y) = \mathrm{Mor}(X, Y)$ whose members are called morphisms in $\mathcal{C}$ from X to Y, which we might in some cases denote by $\mathrm{Hom}(X, Y)$; and
- For each X, Y, Z in $\mathrm{Ob}\,\mathcal{C}$, a map
$$\mathrm{Mor}(Y, Z) \times \mathrm{Mor}(X, Y) \to \mathrm{Mor}(X, Z),$$
referred to as a 'law of composition', denoted $(g, f) \mapsto g \circ f$,

subject to the following properties:

- $\mathrm{Mor}(X, Y)$ and $\mathrm{Mor}(X', Y')$ are disjoint unless $X = X'$ and $Y = Y'$, in which case they are equal;
- *Identity morphisms:* For all $X \in \mathrm{Ob}\,\mathcal{C}$, $\exists\ \mathrm{id}_X \in \mathrm{Mor}(X, X)$ such that for all $Y \in \mathrm{Ob}\,\mathcal{C}$, $f \in \mathrm{Mor}(X, Y)$ and $g \in \mathrm{Mor}(Y, X)$, we have $f \circ \mathrm{id}_X = f$ and $\mathrm{id}_X \circ g = g$.
- *Associativity of composition.* If $f \in \mathrm{Mor}(X, Y), g \in \mathrm{Mor}(Y, Z)$ and $h \in \mathrm{Mor}(Z, W)$ we have
$$(h \circ g) \circ f = h \circ (g \circ f)$$
inside $\mathrm{Mor}(X, W)$.

Remark 2.11. (i) This is not quite a definition because we have not defined what a 'collection' means. It may not be a set: for instance, we will consider categories whose objects are sets, so its 'collection' of objects is the collection of sets, which cannot be a set by Russell's paradox. In this course, we will not worry about such set-theoretic issues, though occasionally we might make remarks about such. Instead, we will just use our usual set-theoretic intuition to think of these 'collections'.

(ii) There are categories $\mathcal{C}$ where $\mathrm{Ob}\,\mathcal{C}$ forms a set, as does the collection of all its morphisms (between varying objects): those are called small categories.

(iii) It is much more common to find categories $\mathcal{C}$ where, for each $X, Y \in \mathrm{Ob}\,\mathcal{C}$, $\mathrm{Mor}(X, Y)$ is a set. These are called 'locally small'. We will almost always only consider locally small categories in this course.

(iv) We just said '$\exists\ \mathrm{id}_X$', not 'we are given id_X': this is because id_X is uniquely determined: if id'_X is another candidate, $\mathrm{id}'_X = \mathrm{id}_X \circ \mathrm{id}'_X = \mathrm{id}_X$.

While talking about morphisms, we will adapt various terminology related to functions without further comment: e.g., We might talk of a morphism f from X to Y and write $f : X \to Y$ instead of saying $f \in \mathrm{Mor}(X, Y)$, id_X will be referred to as the identity morphism from X to X, and for any $f \in \mathrm{Mor}(X, Y)$ we will refer to X as the source or the domain of f and Y as the codomain or the target of f. We might even refer to $f : X \to Y$ as a 'map' from X to Y. The elements of $\mathrm{Mor}(X, X)$ will be referred to as endomorphisms of X.

Example 2.12. (i) The category *Set*: Ob *Sets* is the collection of sets, $\mathrm{Mor}(X, Y)$ is the set of functions $X \to Y$, and where composition is the usual composition of functions. We will refer to this category as the 'category of sets and functions (between sets)', since the composition is understood, or even as just 'the category of sets', when both morphisms and their composition rules are understood. In what follows, we will usually omit describing the composition, and sometimes the morphisms too, but in each case what we omit will be understood from the context.

(ii) *Grp*, the category of groups and group homomorphisms.

(iii) The category whose objects are the groups, and where $\mathrm{Mor}(G, H)$ is the set of equivalence classes of maps $G \to H$, where $f_1 \sim f_2$ if there exists $h \in H$ such that $f_1 = \mathrm{Int}\, h \circ f_2$, where $\mathrm{Int}\, h =$ conjugation by h. Composition is induced by the obvious one: check that it is well-defined.

(iv) *AbGrp*, abelian groups and group homomorphisms.

(v) Top, topological spaces and continuous maps.

(vi) $HTop$, Topological spaces and homotopy classes of continuous maps between them: the well-definedness of composition involves checking, e.g., that if $f_1, f_2 : X \to Y$ are homotopic to each other and $g_1, g_2 : Y \to Z$ are homotopic to each other, then $g_1 \circ f_1, g_2 \circ f_2 : X \to Z$ are homotopic to each other.

(vii) Man, manifolds and smooth maps.

(viii) $Ring$, rings and ring homomorphisms.

(ix) For a commutative ring R, recall that a commutative R-algebra is a commutative ring S together with a ring homomorphism $\iota : R \to S$. Then we have the category of commutative R-algebras: its objects are commutative R-algebras, and the morphisms between two objects (S_1, ι_1) and (S_2, ι_2) are ring homomorpisms $f : S_1 \to S_2$ fitting into a commutative diagram

$$\begin{array}{ccc} S_1 & \xrightarrow{\quad f \quad} & S_2 \ . \\ & {\scriptstyle \iota_1} \nwarrow \quad \nearrow {\scriptstyle \iota_2} & \\ & R & \end{array}$$

(x) $R\text{-}Mod$ (resp., $Mod\text{-}R$) for a not necessarily commutative ring R, the category of left R-modules (resp., right R-modules) and R-module homomorphisms. Note that $\mathbb{Z}\text{-}Mod$ can be identified with $AbGrp$.

(xi) $Vec_k := k\text{-}Mod$, when k is a field, so this is the category of vector spaces over k and k-linear transformations.

(xii) $Ban_{\mathbb{R}}$ (resp., $Ban_{\mathbb{C}}$), Banach spaces over $\mathbb{R}$ (resp., $\mathbb{C}$) and bounded linear maps.

(xiii) Given a group G, the category of G-sets, i.e., sets X together with an action of G, where $\mathrm{Mor}(X, Y)$ is the set of maps $X \to Y$ respecting the G-actions.

(xiv) The category of pairs (G, X) where G is a group acting on a set X; $\mathrm{Mor}((G, X), (H, Y))$ consists of all pairs consisting of a homomorphism $G \to H$ and a function $X \to Y$ with the obvious compatibility: if the former maps g to h and the latter x to y, the latter also maps $g \cdot x$ to $h \cdot y$.

(xv) Pairs (V, T) consisting of a vector space V over a given field k, and a k-linear transformation $T : V \to V$, with

$$\mathrm{Mor}((V, T), (W, U)) = \{f : V \to W \mid f \circ T = U \circ f\}.$$

(xvi) Open subsets of $\mathbb{C}^n$ and holomorphic maps between them.

(xvii) Given a ring R and a group G, the category whose objects are R-modules equipped with a G-action by R-module automorphisms, and whose morphisms are morphisms of R-modules that respect the G-action. If $R = k$ is a field, this is by definition the category $Rep_k G$ of representations of G on k-vector spaces.

Example 2.13. If G is a group, define $*_G$ to be the category such that $\mathrm{Ob}\, *_G = \{*\}$ is a singleton set, $\mathrm{Mor}(*, *) = G$, and composition of morphisms is multiplication in G.

Definition 2.14. (i) In a category $\mathcal{C}$, $f \in \mathrm{Mor}(X, Y)$ is said to be an isomorphism from X to Y if there exists $g \in \mathrm{Mor}(Y, X)$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. If

such an f exists we say that X and Y are isomorphic. Isomorphisms $X \to X$ will be referred to as automorphisms of X, and the collection of these will be denoted by $\mathrm{Aut}(X)$.

(ii) A subcategory $\mathcal{C}'$ of $\mathcal{C}$ is a category $\mathcal{C}'$ such that $\mathrm{Ob}\,\mathcal{C}' \subset \mathrm{Ob}\,\mathcal{C}$, such that for all $X, Y \in \mathrm{Ob}\,\mathcal{C}' \subset \mathrm{Ob}\,\mathcal{C}$, we have $\mathrm{Mor}_{\mathcal{C}'}(X,Y) \subset \mathrm{Mor}_{\mathcal{C}}(X,Y)$, and such that the identity morphisms id_X as well as the compositions in $\mathcal{C}'$ are compatible with those in $\mathcal{C}$.

(iii) If $\mathcal{C}$ is a category, then $\mathcal{C}^{op}$ is the category such that $\mathrm{Ob}\,\mathcal{C}^{op} = \mathrm{Ob}\,\mathcal{C}$, and such that for all $X, Y \in \mathrm{Ob}\,\mathcal{C}$, $\mathrm{Mor}_{\mathcal{C}^{op}}(X,Y) = \mathrm{Mor}_{\mathcal{C}}(Y,X)$, where $g \circ f : X \xrightarrow{f} Y \xrightarrow{g} Z$ in $\mathcal{C}^{op}$ is $Z \xrightarrow{g} Y \xrightarrow{f} X$ in $\mathcal{C}$.

(iv) $\mathcal{C}$ is said to be a groupoid if every morphism in $\mathcal{C}$ is an isomorphism.

Example 2.15. In Grp an isomorphism is an isomorphism of groups, in $Ring$ an isomorphism of rings, in R-Mod an R-module isomorphism, in Top a homeomorphism, in $HTop$ a homotopy equivalence, and in Man a diffeomorphism.

Example 2.16. (i) $*_G$ is clearly a groupoid.

(ii) The category whose objects are all the vector spaces over a field k, but where $\mathrm{Mor}(V,W)$ is simply the set of isomorphisms $V \to W$, is also a category, and is a groupoid. Similarly with groups, rings or any other category.

(iii) If X is a topological space, we can define the category $\mathcal{C}$ with $\mathrm{Ob}\,\mathcal{C} = X$, and where for $x, y \in X = \mathrm{Ob}\,\mathcal{C}$, $\mathrm{Mor}(x,y)$ = the homotopy classes of paths from x to y, i.e., continuous maps $f : [0,1] \to X$ with $f(0) = x$ and $f(1) = y$. Define $g \circ f$ by

$$(g \circ f)(x) = \begin{cases} f(2x), & \text{if } x \in [0, 1/2], \text{ and} \\ g(2x-1), & \text{if } x \in [1/2, 1] \end{cases}$$

(check that it is well-defined and continuous). This is a groupoid (check), called the fundamental groupoid of X: the inverse of $f : [0,1] \to X$ is the reverse path, $g : [0,1] \to X$ such that $g(t) = f(1-t)$ for $0 \leqslant t \leqslant 1$.

2.5. **Functors.** If I understand it right, categories were originally designed to understand (what turned out to be) functors:

Definition 2.17. Let $\mathcal{C}, \mathcal{D}$ be categories. A functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ consists of the following data:

(i) For each $A \in \mathrm{Ob}\,\mathcal{C}$, an object $F(A) \in \mathrm{Ob}\,\mathcal{D}$; and very importantly also

(ii) For all $f : X \to Y$ in $\mathcal{C}$, a morphism $F(f) : F(X) \to F(Y)$ in $\mathcal{D}$, subject to the properties $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ and $F(f \circ g) = F(f) \circ F(g)$.

Let us emphasize that a functor should be defined both at the level of objects and at the level of morphisms, though sometimes one may specify it just at the level of objects when its definition at the level of morphisms is understood.

A functor $\mathcal{C}^{op} \rightsquigarrow \mathcal{D}$ is also referred to sometimes as a *contravariant* functor from $\mathcal{C}$ to $\mathcal{D}$.

Note that functors between categories can be composed.

Example 2.18. (i) We have 'forgetful functors' $Forget : Grp \rightsquigarrow Set$, $Forget : R\text{-}Mod \rightsquigarrow AbGrp$, $Forget : R\text{-}Mod \rightsquigarrow Set$, $Forget : Rep_k G \rightsquigarrow Vect_k$. For instance, $Forget : Grp \rightsquigarrow Set$ assigns to each group its underlying set, and assigns to each group homomorphism $G \to H$ the same map viewed as a map of sets.

(ii) $\pi_0 : Top \rightsquigarrow Set$, assigns to each topological space its set of connected components $\pi_0(X)$, and to each continuous map $f : X \to Y$ of topological spaces the induced map $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$ of connected components: it is well-defined since the image of a connected component of X under the continuous map f is connected and hence contained in a connected component of Y.

(iii) However, we don't have a functor $\pi_1 : Top \rightsquigarrow Grp$: π_1 is not assigned to a topological space X, but to a *pointed topological space* or a *based topological space* (X, x), where X is a topological space and $x \in X$ is a point. There is a category of pointed topological spaces, say $\widetilde{Top}$, where $\mathrm{Mor}((X, x), (Y, y))$ is the set of continuous maps $f : X \to Y$ such that $f(x) = y$. Such a map f uniquely determines a group homomorphism $\pi_1(f) : \pi_1(X, x) \to \pi_1(Y, y)$. This respects composition and identity morphisms, so assigning to $\pi_1(X, x)$ to (X, x) and $\pi_1(f) : \pi_1(X, x) \to \pi_1(Y, y)$ to $f : (X, x) \to (Y, y)$, defines a functor $\pi_1 : \widetilde{Top} \to Grp$.

One way to describe this is the following: if X is a topological space, then without the choice of a base-point we can define π_1 up to an isomorphism but not up to a unique isomorphism. This issue makes it non-functorial at the level of the category Top, because a functor is defined at the level of morphisms as well, not just objects.

(iv) A functor $F : *_G \rightsquigarrow Set$ is simply a set with an action of G: $X := F(*)$ is a set, for all $g \in G = \mathrm{Mor}(*, *)$, F gives $F(g) \in \mathrm{Mor}_{Set}(F(*), F(*)) = \{\text{Maps } X \to X\}$, and the rules $F(g) \circ F(h) = F(g \circ h)$ and $F(\mathrm{id}_*) = \mathrm{id}_{F(*)}$ translate to $F(gh) = F(g)F(h)$ and that $F(\mathrm{id}_*)$ is the identity map $X \to X$. Thus, $g \mapsto F(g)$ is a group homomorphism $G \to \mathrm{Bij}(X, X)$, which is the same as giving an action of G on X: $g \cdot x = F(g)(x)$.

(v) By the same reasoning, a functor $F : *_G \rightsquigarrow Vec_k$ is simply a representation of G on a k-vector space. More generally, a functor $F : *_G \to \mathcal{C}$ can be thought of as an object of $\mathcal{C}$ with an action of G.

(vi) Any group homomorphism $G \to H$ induces a functor $*_G \rightsquigarrow *_H$.

Lemma 2.19. *Let $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ be a functor. If $X, Y \in \mathrm{Ob}\,\mathcal{C}$ are isomorphic, then so are $F(X), F(Y) \in \mathrm{Ob}\,\mathcal{D}$.*

Proof. If $f : X \to Y$ and $g : Y \to X$ are such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$, then $F(f) : F(X) \to F(Y)$ and $F(g) : F(Y) \to F(X)$ are such that $F(g) \circ F(f) : F(X) \to F(X)$ equals $F(g \circ f) = F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$, and similarly $F(f) \circ F(g) = \mathrm{id}_{F(Y)}$. This shows that $F(f) : F(X) \to F(Y)$ is an isomorphism with inverse $F(g) : F(Y) \to F(X)$. □

A consequence of the above lemma: Let X and Y be homeomorphic path connected topological spaces, with $f : X \to Y$ a homeomorphism. Then for any $x \in X$, letting $y := f(x)$,

(X, x) and (Y, y) are isomorphic in the category $\widetilde{Top}$ of pointed topological spaces. Thus, by the above lemma, $\pi_1(X, x) \cong \pi_1(Y, y)$. In other words, we can show two path connected topological spaces to be non-homeomorphic, if we show that their fundamental groups are not isomorphic: e.g.: $\mathbb{R}^2 \backslash \{0\}$ and $\mathbb{R}^2$.

In your topology course, you will see functors:

$$H_i : Top \rightsquigarrow AbGrp, \qquad H^i : Top^{op} \rightsquigarrow AbGrp,$$

for each integer $i \geqslant 0$. This can be sometimes be used to show that two given topological spaces are not homeomorphic. For instance, for each $n \geqslant 1$, we have:

$$H_i(S^n) \cong \begin{cases} \mathbb{Z}, & \text{if } i = 0 \text{ or } n, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Thus, if $m \neq n$ and $m, n \geqslant 1$, then by Lemma 2.19 we have that S^m and S^n are not homeomorphic to each other, since $H_n(S^m) = 0 \ncong \mathbb{Z} \cong H_n(S^m)$.

2.6. **Full, faithful and essentially surjective functors.**

Definition 2.20. A functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ is said to be

(i) faithful (resp., full; resp., fully faithful), if for all $X, Y \in \operatorname{Ob}\mathcal{C}$, the map $\operatorname{Mor}(X, Y) \to \operatorname{Mor}(F(X), F(Y))$ given by $f \mapsto F(f)$ is injective (resp., surjective; resp., bijective).
(ii) essentially surjective, if for all $A \in \operatorname{Ob}\mathcal{D}$, there exists $X \in \operatorname{Ob}\mathcal{C}$ such that $F(X)$ is isomorphic to A in the category $\mathcal{D}$ (we are not requiring that A itself is of the form $F(X)$, it just needs to be isomorphic to something of that form).
(iii) an equivalence of categories, if it is fully faithful and essentially surjective.

Example 2.21. (i) $Forget : Grp \rightsquigarrow Sets$ and $Forget : R\text{-}Mod \rightsquigarrow AbGrp \rightsquigarrow Set$ and $Forget : Top \rightsquigarrow Set$ are all faithful, but none of them is full. The obvious 'inclusion functor' $AbGrp \rightsquigarrow Grp$ is fully faithful.

(ii) If $G \to H$ is a group homomorphism, the functor $*_G \rightsquigarrow *_H$ discussed in Example 2.18(vi) is faithful (resp., full; resp., fully faithful) if and only if $G \to H$ is injective (resp., surjective; resp., bijective).

(iii) Consider the category Vec_k^{fd} of finite dimensional k-vector spaces and k-linear transformations, and its full subcategory $\mathcal{C}$ consisting of vector spaces of the form k^n for some $n \in \mathbb{Z}_{\geqslant 0}$: this means that the members of $\operatorname{Ob}\mathcal{C}$ are simply the k-vector spaces of the form k^n, and that $\operatorname{Mor}_{\mathcal{C}}(X, Y) = \operatorname{Mor}_{Vec_k^{fd}}(X, Y)$ for all $X, Y \in \operatorname{Ob}\mathcal{C}$. Then, by definition, the inclusion functor $\mathcal{C} \rightsquigarrow Vec_k^{fd}$ is fully faithful. Since every finite dimensional k-vector space is isomorphic to some k^n, it is also essentially surjective, and hence is an equivalence of categories. Note that $\mathcal{C}$ is small, while Vec_k^{fd} is not.

(iv) Vec_k^{fd} is equivalent to $(Vec_k^{fd})^{op}$, by the functor that takes V in $\operatorname{Ob} Vec_k^{fd}$ to its dual $V^\vee := \operatorname{Hom}_k(V, k)$, and each linear map $T : V \to W$ to the transpose (or "pull-back under T") map ${}^tT : W^\vee \to V^\vee$, thought of as an element of $\operatorname{Mor}_{(Vec_k^{fd})^{op}}(V^\vee, W^\vee)$.

(v) Later, we will hopefully see that for any natural number $n \geqslant 1$, Vec_k^{fd} is equivalent to "$\mathrm{M}_n(k)$-$Mod^{f.g.}$", by a functor that, at the level of objects, takes V to "$V \otimes_k k^n$".

Fully faithful functors satisfy the following stronger ("if and only if") version of Lemma 2.19:

Lemma 2.22. *Let $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ be a fully faithful functor. Then two objects $X, Y \in \mathrm{Ob}\,\mathcal{C}$ are isomorphic if and only if $F(X), F(Y) \in \mathrm{Ob}\,\mathcal{D}$ are.*

Proof. Easy exercise. This property of F is referred to as F being conservative. □

2.7. **Natural transformations.**

Definition 2.23. (i) Let $F, G : \mathcal{C} \rightsquigarrow \mathcal{D}$ be functors. A natural transformation ϕ from F to G is a collection of morphisms in $\mathcal{D}$ indexed by $\mathrm{Ob}\,\mathcal{C}$,

$$\phi = (\phi_X : F(X) \to G(X))_{X \in \mathrm{Ob}\,\mathcal{C}}$$

(i.e., each ϕ_X lies in $\mathrm{Mor}_{\mathcal{D}}(F(X), G(X))$), respecting morphisms in the sense that for all $f : X \to Y$ in $\mathcal{C}$, the following diagram commutes:

$$\begin{array}{ccc} F(X) & \xrightarrow{\phi_X} & G(X) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(Y) & \xrightarrow{\phi_Y} & G(Y) \end{array} .$$

(ii) We say that ϕ as above is a natural isomorphism if it has an inverse natural transformation, i.e., a natural transformation ψ from G to F such that for all $X \in \mathrm{Ob}\,\mathcal{C}$,

$$\psi_X \circ \phi_X : F(X) \overset{\phi_X}{\to} G(X) \overset{\psi_X}{\to} F(X) \qquad \text{and } \phi_X \circ \psi_X : G(X) \overset{\psi_X}{\to} F(X) \overset{\phi_X}{\to} G(X)$$

are identity morphisms, namely $\mathrm{id}_{F(X)}$ and $\mathrm{id}_{G(X)}$. In other words, the composite natural transformations $\phi \circ \psi$ and $\psi \circ \phi$ are the identity functor $\mathcal{D} \to \mathcal{D}$.

(iii) Given categories $\mathcal{C}$ and $\mathcal{D}$, we have a category $\mathrm{Fun}(\mathcal{C}, \mathcal{D})$, whose objects are the functors from $\mathcal{C}$ to $\mathcal{D}$, and where the morphisms between two functors F and G are the natural transformations ϕ from F to G (composition is understood to be the composition of natural transformations).

Example 2.24. Recall Examples 2.18 (iv) and (v): expanding on the reasoning there, it follows that the category $\mathrm{Fun}(*_G, Set)$ is the category of sets with a G-action (Example 2.12(xiii)), and the category $\mathrm{Fun}(*_G, Vec_k)$ is $Rep_k G$.

Remark 2.25. One can show that the functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ is an equivalence of categories if and only if there exists a functor $G : \mathcal{D} \rightsquigarrow \mathcal{C}$ such that $G \circ F$ is naturally isomorphic (in the category $\mathrm{Fun}(\mathcal{C}, \mathcal{C})$) to the identity functor $\mathcal{C} \rightsquigarrow \mathcal{C}$, and such that $F \circ G$ is naturally isomorphic to the identity functor of $\mathcal{D} \rightsquigarrow \mathcal{D}$. For more details, see Arvind's notes; this result uses a form of axiom of choice that applies to classes which may not be sets.

Note that $G \circ F$ and $F \circ G$ are not required to be identity functors at all: that would make the definition too restrictive to be useful; there is a notion of 'isomorphism of categories', which is not nearly as useful as equivalence of categories.

3. Lecture 3 — Yoneda lemma, more on categories and functors

3.1. **Preliminary comments.** Remember that we ignore set-theoretic difficulties.

Correction: For $\mathcal{C}$ to be a small category, not only do we require that $\mathrm{Ob}\,\mathcal{C}$ is a set, but also that for all $X, Y \in \mathrm{Ob}\,\mathcal{C}$, $\mathrm{Mor}_{\mathcal{C}}(X, Y)$ is a set. Varying this over X, Y belonging to the set $\mathrm{Ob}\,\mathcal{C} \times \mathrm{Ob}\,\mathcal{C}$, the collection of all morphisms in $\mathcal{C}$ also forms a set. (Last time, I carelessly copied the definition from a source which was already assuming that its categories were locally small, so for that source the condition that the objects formed a set was enough.)

Today: by 'category', we will mean a 'locally small category': each $\mathrm{Mor}_{\mathcal{C}}(X, Y)$ is a set.

Recall that given two categories $\mathcal{C}$ and $\mathcal{D}$, we have a category $\mathrm{Fun}(\mathcal{C}, \mathcal{D})$ whose objects are functors $F : \mathcal{C} \to \mathcal{D}$, and whose morphisms are natural transformations. Isomorphisms in this category are called 'natural isomorphisms', and are what we refer to when we say 'is naturally isomorphic to'.

Notation 3.1. Henceforth, if $F, G : \mathcal{C} \rightsquigarrow \mathcal{D}$ are functors, we will write $\mathrm{Nat}(F, G)$ for the collection of natural transformation from F to G.

Warning: For ease of communication, I may occasionally refer to certain objects 'universal objects'; that is not standard terminology and should not be used in formal mathematical writing.

3.2. **The Yoneda embeddings.** *Very informal motivation.* From a category theoretic perspective, rather than looking into structures that constitute an object (e.g., addition and scalar multiplication in a vector space), one tries to glean information about objects on the basis of their morphisms to or from other objects. An analogy I remember hearing from Professor Nitin Nitsure is: we can't look inside anyone's head, but we can get some information about what they are thinking based on what they talk with other people!

Thus, to get information about $X \in \mathrm{Ob}\,\mathcal{C}$ we might use:

Definition 3.2. Let $X \in \mathrm{Ob}\,\mathcal{C}$, where $\mathcal{C}$ is a (locally small by convention) category.

(i) By $h_X := \mathrm{Mor}_{\mathcal{C}}(X, -)$, we denote the functor $\mathcal{C} \rightsquigarrow \mathit{Set}$ defined as follows:
- *At the level of objects:* For $Y \in \mathrm{Ob}\,\mathcal{C}$, $h_X(Y) = \mathrm{Mor}_{\mathcal{C}}(X, Y)$ (a set since $\mathcal{C}$ is locally small).
- *At the level of morphisms:* If $f : Y \to Z$ is a morphism in $\mathcal{C}$, then
$$h_X(f) : h_X(Y) = \mathrm{Mor}_{\mathcal{C}}(X, Y) \overset{(Y \to Z)\circ -}{\to} \mathrm{Mor}_{\mathcal{C}}(X, Z) = h_X(Z)$$
is given by post-composition in $\mathcal{C}$ with $Y \to Z$.

(ii) Analogously, we have the functor $h^X := \mathrm{Mor}_{\mathcal{C}}(-, X)$, a functor $\mathcal{C}^{op} \rightsquigarrow \mathit{Set}$: for $Y \in \mathrm{Ob}\,\mathcal{C}$ and $f : Y \to Z$ in $\mathcal{C}^{op}$, i.e., $Z \to Y$ in $\mathcal{C}$,
$$h^X(Y) = \mathrm{Mor}_{\mathcal{C}}(Y, X),$$
$$h^X(f) : h^X(Y) = \mathrm{Mor}_{\mathcal{C}}(Y, X) \overset{-\circ(Z \to Y)}{\to} \mathrm{Mor}_{\mathcal{C}}(Z, X) = h^X(Z).$$

Definition 3.3. Let $\mathcal{C}$ be a category. The category of presheaves on $\mathcal{C}$ is the category $\mathrm{Presh}(\mathcal{C}) := \mathrm{Fun}(\mathcal{C}^{op}, Set)$, i.e., the category of contravariant functors from $\mathcal{C}$ to the category Set of sets. We will also be interested in $\mathrm{Presh}(\mathcal{C}^{op}) = \mathrm{Fun}(\mathcal{C}, Set)$.

As we will see, these categories contain a copy of $\mathcal{C}$ or $\mathcal{C}^{op}$ via the h^X or the h_X: they are larger categories (e.g., if $\mathcal{C}$ has just one object and one morphism, $\mathrm{Presh}(\mathcal{C})$ identifies with the category Set of sets, which is not a small category), closed under certain operations that $\mathcal{C}$ or $\mathcal{C}^{op}$ may not be, as we hope the discussion on products and coproducts below will show (see Remarks 3.21 and 3.25).

Example 3.4. If X is a topological space, one can consider the category $\mathcal{C}$ of open subsets of X, where the morphisms are inclusions. In this case, the objects of $\mathrm{Presh}(\mathcal{C})$ are simply the presheaves of sets on $\mathcal{C}$ in the usual sense.

More examples will be found in the following definitions of Yoneda embeddings.

Definition 3.5. (i) We upgrade $X \rightsquigarrow h_X$, defined so far at the level of objects, to a functor $h_\bullet : \mathcal{C}^{op} \to \mathrm{Fun}(\mathcal{C}, Set)$: if $X \to Y$ is a morphism in $\mathcal{C}$, corresponding to $f : Y \to X$ in $\mathcal{C}^{op}$, then

$$h_\bullet(f) : h_Y = \mathrm{Mor}_{\mathcal{C}}(Y, -) \overset{-\circ X \to Y}{\to} \mathrm{Mor}_{\mathcal{C}}(X, -) = h_X$$

is the natural transformation given by pre-composition with $X \to Y$.

(ii) Similarly, $X \rightsquigarrow h^X$ can also be upgraded to a functor $h^\bullet : \mathcal{C} \to \mathrm{Presh}(\mathcal{C}) = \mathrm{Fun}(\mathcal{C}^{op}, Set)$: if $f : X \to Y$ is a morphism in $\mathcal{C}$, its image $h^\bullet(f) \in \mathrm{Nat}(h^X, h^Y)$ is the natural transformation $\mathrm{Mor}_{\mathcal{C}}(-, X) \to \mathrm{Mor}_{\mathcal{C}}(-, Y)$ given by post-composition with $X \to Y$.

(iii) The functor $h_\bullet : \mathcal{C}^{op} \to \mathrm{Fun}(\mathcal{C}, Set)$ given by $X \rightsquigarrow h_X$, and the functor $h^\bullet : \mathcal{C} \to \mathrm{Presh}(\mathcal{C}) = \mathrm{Fun}(\mathcal{C}^{op}, Set)$ given by $X \rightsquigarrow h^X$ [3] are called Yoneda embeddings. e.g., the latter may be called the Yoneda embedding of $\mathcal{C}$ in its presheaf category.

Remark 3.6. Why 'embedding'? A functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ is an embedding if it is both injective on objects and faithful (e.g., a quasi-inverse to the inclusion of vector spaces of the form k^n in all finite dimensional k-vector spaces is faithful but not injective on objects). It is easy to check that the Yoneda embeddings are embeddings; the injectivity on objects comes from the definitional fact that $\mathrm{Mor}_{\mathcal{C}}(X, Y)$ and $\mathrm{Mor}_{\mathcal{C}}(X', Y')$ are disjoint unless $X = X'$ and $Y = Y'$. Below, we will see the Yoneda lemma, which says that these are not just embeddings, but also *fully* faithful.

Example 3.7. This is an important special case of the h^X of Definition 3.2(ii) above. Let $\mathcal{C} = (k\text{-}alg^{fg})^{op}$ be the opposite category of the category of finitely generated k-algebras (and k-algebra homomorphisms between them): here in in the rest of this lecture, a k-algebra will be understood to be commutative. Then $\mathcal{C}$ is referred to as the category of affine algebraic schemes over k. Let $X \in \mathrm{Ob}\,\mathcal{C} = \mathrm{Ob}\,\mathcal{C}^{op}$ be the finitely generated k-algebra $k[x_1, \ldots, x_n]/(f_1, \ldots, f_m)$ (with k understood to map to it in the obvious way). Then:

[3] Recall, this is abuse of notation: strictly speaking, the functors include their definition at the level of morphisms too, which are suppressed here for brevity.

- What is $h^X(Y)$? Since Y is a f.g. k-algebra, let us write $k \to R$ for it.
$$\begin{aligned} h^X(Y) &= \mathrm{Mor}_{\mathcal{C}}(Y, X) = \mathrm{Mor}_{k-alg}(k[x_1, \ldots, x_n]/(f_1, \ldots, f_m), R) \\ &= \{(a_1, \ldots, a_n) \in R^n \mid \forall\, 1 \leqslant j \leqslant m, f_j(a_1, \ldots, a_n) = 0\}. \end{aligned}$$
Thus, for each $Y = (k \to R)$, $h^X(Y)$ is simply the set of solutions in R^n to the system $f_1 = \cdots = f_m = 0$ of equations in n variables over k, viewed via $k \to R$ as a system of equations over R.
- Similarly, if $Z \to Y$ is a morphism in $\mathcal{C}$, corresponding to a morphism $R \to S$ of k-algebras (i.e., in $\mathcal{C}^{op}$, so Y corresponds to R and Z to S), then then $h^X(Y) \to h^X(Z)$ is the obvious map

$$\{\text{Solutions of } f_1 = \cdots = f_m = 0 \text{ in } R^n\} \to \{\text{Solutions of } f_1 = \cdots = f_m = 0 \text{ in } S^n\},$$

given by applying $R \to S$: note that this makes sense since $R \to S$ is a k-algebra homomorphism (e.g., if it takes r to s then it takes $2r^2 + 3r$ to $2s^2 + 3s$), and $f_1, \ldots, f_m \in k[x_1, \ldots, x_n]$.

Definition 3.8. A presheaf $F \in \mathrm{Fun}(\mathcal{C}^{op}, Set)$ is said to be representable if it is naturally isomorphic to h^X, for some $X \in \mathrm{Ob}\,\mathcal{C}$. Similarly, $F \in \mathrm{Fun}(\mathcal{C}, Set)$ is said to be corepresentable if it is naturally isomorphic to h_X, for some $X \in \mathrm{Ob}\,\mathcal{C}$.

Example 3.9. As in Example 3.7, let $\mathcal{C} = (k\text{-}alg^{fg})^{op}$. Then an element of $\mathrm{Presh}(\mathcal{C})$ can be thought of as a functor $F : k\text{-}alg^{fg} \to Set$, say $R \rightsquigarrow F(R)$ (at the level of objects). To say that this functor is representable, means:

- Informally, it means that $F(R)$, as R-varies over finitely generated k-algebras, is (bijective, functorially in R with) the set of solutions in R^n to some fixed set of equations $f_1 = \cdots = f_m = 0$ over k.
- Formally, of course, it means that there is some finitely generated k-algebra A with the property that for each finitely generated k-algebra R, we have an identification $\phi_R : F(R) \overset{\text{bij.}}{\to} \mathrm{Hom}_{k\text{-}alg}(A, R)$ for each f.g. k-algebra R, which is furthermore functorial in R – i.e., for any k-algebra homomorphism $f : R \to S$, the diagram below commutes:
$$\begin{array}{ccc} F(R) & \xrightarrow{\phi_R} & \mathrm{Hom}_{k\text{-}alg}(A, R) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle f\circ -} \\ F(S) & \xrightarrow{\phi_S} & \mathrm{Hom}_{k\text{-}alg}(A, S). \end{array}$$

How to translate from the informal to the formal perspective? If the equations are $f_1 = \cdots = f_m = 0$ in n variables $x_1, \ldots, x_n$, then $A = k[x_1, \ldots, x_n]/(f_1, \ldots, f_m)$. [4]

Example 3.10. The forgetful functor $AbGrp \rightsquigarrow Set$ is corepresented by $\mathbb{Z}$, since for any abelian group A, $\mathrm{Mor}_{AbGrp}(\mathbb{Z}, A)$ identifies with A via $\varphi \mapsto \varphi(1)$. Similarly the forgetful functor $R\text{-}Mod \rightsquigarrow Set$ is corepresented by R.

[4] In fact any finitely generated k-algebra A accepts a surjection from $k[x_1, \ldots, x_n]$, by whose Noetherianness (because of the Hilbert basis theorem) A is isomorphic to some $k[x_1, \ldots, x_n]/(f_1, \ldots, f_m)$.

3.3. **Yoneda lemma.**

Lemma 3.11 (Yoneda)**.** *The Yoneda embeddings $h_\bullet$ and $h^\bullet$ are fully faithful. In other words, letting $\mathcal{C}$ be a category:*

(i) $\forall\ X, Y \in \mathrm{Ob}\,\mathcal{C}$, $u \mapsto h_\bullet(u)$ *defines a bijection* $\mathrm{Mor}_{\mathcal{C}}(Y, X) = \mathrm{Mor}_{\mathcal{C}^{op}}(X, Y) \to \mathrm{Nat}(h_X, h_Y)$*; and*

(ii) $\forall\, X, Y \in \mathrm{Ob}\,\mathcal{C}$, $u \mapsto h^\bullet(u)$ *defines a bijection* $\mathrm{Mor}_{\mathcal{C}}(X, Y) \to \mathrm{Nat}(h^X, h^Y)$.

This lemma turns out to be a special case of the following version of it – you might possibly prefer to defer reading the precise description of the bijections in the statement of the lemma until you read the 'Idea of the proof' below:

Lemma 3.12 (Yoneda)**.** *(i) For any $X \in \mathrm{Ob}\,\mathcal{C}$ and any functor $F : \mathcal{C} \rightsquigarrow Set$, the map*

$$\mathrm{Nat}(h_X, F) \to F(X),$$

given by

$$\eta \mapsto \underbrace{\eta_X}_{\mathrm{Mor}(X,X)\to F(X)}\ (\ \underbrace{\mathrm{id}_X}_{\in \mathrm{Mor}(X,X)}\ ),$$

is a bijection, with inverse

$$u \mapsto \eta_u = (\eta_{u,Y})_{Y\in \mathrm{Ob}\,\mathcal{C}}, \quad \text{where } \eta_{u,Y} : h_X(Y) \to F(Y) \text{ sends } f : X \to Y \text{ to } \underbrace{F(f)}_{F(X)\to F(Y)}\ (\underbrace{u}_{\in F(X)}).$$

(ii) For any $X \in \mathrm{Ob}\,\mathcal{C}$ and any functor $G : \mathcal{C}^{op} \rightsquigarrow Set$, the map

$$\mathrm{Nat}(h^X, G) \to G(X)$$

given by

$$\eta \mapsto \underbrace{\eta_X}_{\mathrm{Mor}(X,X)\to G(X)}\ (\ \underbrace{\mathrm{id}_X}_{\in \mathrm{Mor}(X,X)}\ )$$

is a bijection, with inverse

$$u \mapsto \eta_u = (\eta_{u,Y})_{Y\in \mathrm{Ob}\,\mathcal{C}}, \quad \text{where } \eta_{u,Y} : h^X(Y) \to G(Y) \text{ sends } f : Y \to X \text{ to } \underbrace{G(f)}_{G(X)\to G(Y)}\ (\underbrace{u}_{\in G(X)}).$$

Moreover, the bijection of (i) is "natural" in X and F, while that of (ii) is "natural" in X and G.

Proof. Both the map and its claimed inverse have been given, so you can verify that they are indeed two-sided inverses of each other. The naturality is also easy to verify. So you can do this as an exercise. But we explain the idea below.

Idea of the proof, for (ii). Somehow the point of the proof is the following feature of h^X. There is a 'universal object' in $h^X(X) = \mathrm{Mor}_{\mathcal{C}}(X, X)$, namely, the identity morphism

$u_0 := \mathrm{id}_X$. Here, the 'universality' means the following: any $f \in h^X(Y) = \mathrm{Mor}_{\mathcal{C}}(Y, X)$, is also the composition of $\mathrm{id}_X : X \to X$ with $f : Y \to X$, so

$$\underbrace{h^X(f)}_{\mathrm{Mor}_{\mathcal{C}}(X,X) \to \mathrm{Mor}_{\mathcal{C}}(Y,X)} \quad (\quad \underbrace{\mathrm{id}_X}_{\in \mathrm{Mor}_{\mathcal{C}}(X,X)} \quad) = f.$$

This means that every natural transformation $\eta : h^X \to G$ is entirely determined by where $\eta_X : h^X(X) \to G(X)$ sends id_X:

$$\begin{array}{ccccc} \mathrm{id}_X \in & & h^X(X) & \xrightarrow{\eta_X} & G(X) \qquad \ni u = \eta_X(\mathrm{id}_X) \, . \\ \downarrow & & \downarrow{\scriptstyle h^X(f)} & & \downarrow{\scriptstyle G(f)} \qquad \downarrow \\ f \in & & h^X(Y) & \xrightarrow{\eta_Y} & G(Y) \qquad \ni G(f)(u) \end{array}$$

Make sure you can read both the 'forward' and 'backward' directions of the bijection from the above diagram. □

Proof of Lemma 3.11. We will prove (ii); the proof of (i) will be suitably analogous. Applying Lemma 3.12 with $G := h^Y$, we get a bijection

$$\mathrm{Mor}_{\mathcal{C}}(X, Y) = h^Y(X) = G(X) \to \mathrm{Nat}(h^X, G) = \mathrm{Nat}(h^X, h^Y),$$

given by

$$\mathrm{Mor}_{\mathcal{C}}(X, Y) \ni u \mapsto (h^X(Z) = \mathrm{Mor}_{\mathcal{C}}(Z, X) \ni f \mapsto h^Y(f)(u) \in \mathrm{Mor}_{\mathcal{C}}(Z, Y) = h^Y(Z))_{Z \in \mathrm{Ob}\,\mathcal{C}}.$$

But $h^Y(f)(u)$ is simply the composite of $u : X \to Y$ with $f : Z \to X$, namely $u \circ f$. Thus, the above bijection $\mathrm{Mor}_{\mathcal{C}}(X, Y) \to \mathrm{Nat}(h^X, h^Y)$ simply sends u to post-composition with u, which is by definition $h^\bullet(u)$ (see Definition 3.5(ii)). Thus, $u \mapsto h^\bullet(u)$ is a bijection $\mathrm{Mor}_{\mathcal{C}}(X, Y) \to \mathrm{Nat}(h^X, h^Y)$, as desired. □

Remark 3.13. Here is a rephrasing of the 'naturality assertion' of Lemma 3.12: one has an obvious notion $\mathcal{C} \times \mathcal{D}$ of a product of two categories $\mathcal{C}$ and $\mathcal{D}$, where the objects of $\mathcal{C} \times \mathcal{D}$ are $\mathrm{Ob}\,\mathcal{C} \times \mathrm{Ob}\,\mathcal{D}$, and where

$$\mathrm{Mor}_{\mathcal{C} \times \mathcal{D}}((X_1, Y_1), (X_2, Y_2)) = \mathrm{Mor}_{\mathcal{C}}(X_1, X_2) \times \mathrm{Mor}_{\mathcal{D}}(Y_1, Y_2).$$

Then the naturality assertion (say, for the latter bijection) in Lemma 3.12 says that the collection of bijections $F(X) \to \mathrm{Nat}(h^X, F)$ there form a natural isomorphism between the following functors $\mathcal{C} \times \mathrm{Fun}(\mathcal{C}^{op}, Set) \rightsquigarrow Set$: the evaluation functor sending (X, F) to $F(X)$, and the functor sending (X, F) to $\mathrm{Nat}(h^X, F)$.

Example 3.14. Check that the functor $GL_n/k : k\text{-}alg^{fg} \rightsquigarrow Set$ that sends $(k \to R)$ to $GL_n(R)$ is representable on $(k\text{-}alg^{fg})^{op}$, by $k[y, x_{ij} \mid 1 \leqslant i, j \leqslant n]/(y \cdot \det(x_{ij}) - 1)$. So is the functor $\mathbb{A}^n/k$ that sends R to R^n, by $k[x_1, \ldots, x_n]$. We would like to intuitively think of GL_n/k, which is a functor, as a group: in fact it can be viewed as a functor $GL_n/k : k\text{-}alg^{fg} \rightsquigarrow Grp$, and we would like to think of it as acting on $\mathbb{A}^n/k$, thought of as column vectors, by matrix multiplication. In particular, we would like to consider the orbit map $GL_n \to \mathbb{A}^n/k$ sending $g \mapsto g \cdot {}^t(1, 0, \ldots, 0)$. One way to do this is to define a suitable

map $k[x_1, \ldots, x_n] \to k[y, x_{ij} \mid 1 \leqslant i, j \leqslant n]/(y \cdot \det(x_{ij}) - 1)$. This is not exactly pleasant. But the Yoneda lemma, 3.11, implies that this specification can also be done by specifying the resulting map $GL_n(R) \to R^n$ for each k-algebra R: namely, it sends $(a_{ij})_{1 \leqslant i,j \leqslant n}$ to ${}^t[a_{11} \ldots a_{n1}]$; it is functorial in R.

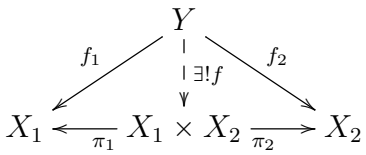
Similarly, we can think of the "action" map $GL_n/k \times \mathbb{A}^n/k \to \mathbb{A}^n/k$ by giving, for each f.g. k-algebra R, the matrix multiplication formulation of $GL_n(R) \times R^n \to R^n$, observing that this is both a group action and functorial in R (and hence a natural transformation). Again by the Yoneda lemma, this gives us a complicated map of rings.
Upshot: Thus, in many contexts, giving an element of $Nat(h^X, h^Y)$ is simpler and more intuitive than giving an element of $\mathrm{Mor}_{\mathcal{C}}(X, Y)$.

Remark 3.15. (Some sort of "universal objects"; warning: this terminology is informal, don't use it in formal writing) Now assume that a presheaf $F \in \mathrm{Fun}(\mathcal{C}^{op}, Set)$ is representable, say it is represented by X. In other words, there exists a natural isomorphism from $\eta : h^X \to F$. By Lemma 3.11, this natural isomorphism is described by an element $u \in F(X)$: in terms of u, $\eta_Y : h^X(Y) = \mathrm{Mor}_{\mathcal{C}}(Y, X) \to F(Y)$ is given by $\eta_Y(f) = F(f)(u)$: in other words, the set $F(Y)$ consists of various possible images of the 'universal object' $u \in F(X)$ under the various $F(f)$, as f varies over $\mathrm{Mor}_{\mathcal{C}}(Y, X)$.

This sort of a 'universal object' is quite common in mathematics. An example can be seen in Remark 3.21 below, shortly below (13). Here is an example of the analogous phenomenon in the context of $h_\bullet$ and corepresentable functors, especially for those of you who have some familiarity with tensor products. Later we will construct a tensor product $M \otimes_R N$ of R-modules M and N using the representability of the functor that sends an R-module L to the set of R-bilinear maps from $M \times N \to L$, and then the 'u' as above will be a bilinear map $M \times N \to M \otimes_R N$. That every element of $F(Y)$ consists of the various $F(f)(u)$ will then correspond to the fact that every bilinear map $M \times N \to L$ will be the composite of some $M \otimes_R N \to L$ (the "f") and $M \times N \to M \otimes_R N$ (the "u").

3.4. **Products.**

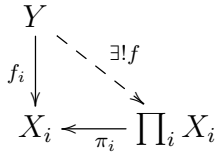
Definition 3.16. (i) Let $X_1, X_2 \in \mathrm{Ob}\,\mathcal{C}$. A product of X_1 and X_2 is a triple (X, π_1, π_2) consisting of an object $X \in \mathcal{C}$, typically denoted $X_1 \times X_2$, and morphisms $\pi_1 : X \to X_1$ and $\pi_2 : X \to X_2$, satisfying the following universal property: For all $Y \in \mathrm{Ob}\,\mathcal{C}$ and every pair of morphisms $f_1 : Y \to X_1$ and $f_2 : Y \to X_2$, there exists a unique morphism $f : Y \to X = X_1 \times X_2$ such that $f_1 = \pi_1 \circ f$ and $f_2 = \pi_2 \circ f$:



In other words, the following map is a bijection:

$$\mathrm{Mor}_{\mathcal{C}}(Y, X) \overset{(\pi_1 \circ -, \pi_2 \circ -)}{\to} \mathrm{Mor}_{\mathcal{C}}(Y, X_1) \times \mathrm{Mor}_{\mathcal{C}}(Y, X_2). \tag{11}$$

(ii) We can similarly define $(X = \prod_{i\in I} X_i, (\pi_i : X \to X_i)_{i\in I})$, a product of a family $(X_i)_{i\in I}$ indexed by some set I:



In other words, the following map is a bijection:

$$\mathrm{Mor}_{\mathcal{C}}(Y, X) \overset{(\pi_i \circ -)_{i\in I}}{\to} \prod_{i\in I} \mathrm{Mor}_{\mathcal{C}}(Y, X_i). \tag{12}$$

(i) is a special case of this, and will be referred to as a 'binary product'.

(iii) We can interpret or extend the definition in (ii) to apply to the case where $I = \varnothing$: An 'empty product' in a category $\mathcal{C}$ is by definition a *terminal object* or a *final object* in the category, by which one means an object $* \in \mathrm{Ob}\,\mathcal{C}$ such that for all $Y \in \mathrm{Ob}\,\mathcal{C}$, $\mathrm{Mor}_{\mathcal{C}}(Y, *)$ is a singleton.

(iv) We say that a category $\mathcal{C}$ has small products if every collection $(X_i)_{i\in I}$ of objects of $\mathcal{C}$ has a product (since I can be empty, this includes the requirement that $\mathcal{C}$ has a terminal object). These are called small products ('small' because I is a set)

A product in a category need not exist, but if it does, it is suitably unique:

Exercise 3.17. Show that in a category $\mathcal{C}$, a product of $(X_i)_{i\in I}$ need not exist, but if it exists, it is 'uniquely unique': if $(X, (\pi_i)_{i\in I})$ and $(X', (\pi'_i)_{i\in I})$ are both products of the X_i, then there exists a *unique* isomorphism $\tau : X \to X'$ such that $\pi_i = \pi'_i \circ \tau$ for each $i \in I$. (For $I = \varnothing$, this is saying that a terminal object of the category, if it exists, is uniquely unique).

The 'uniquely' in the assertion of Exercise 3.17 refers to the fact that the product is not just unique up to an isomorphism, but unique up to a *unique* isomorphism. The uniqueness assertion in the exercise justifies writing $\prod_i X_i$ for the (object underlying the tuple constituting) the product of the X_i (and similarly $X_1 \times X_2$ for a binary product).

Example 3.18. The following categories have arbitrary small products, which coincide with what you already have seen called the products of their objects: $Set, Grp, Top, AbGrp, R\text{-}Mod, Vec_k, Ring$. For instance, if $(X_i)_{i\in I}$ is a family of topological spaces, we take $\prod_{i\in I} X_i$ to be the set-theoretic product of the X_i, given the product topology, and $\pi_j : \prod_{i\in I} X_i \to X_j$ to be the projection onto the j-th factor. Except, we also need to worry about the empty product, namely a terminal object, which exists in each of these cases: a singleton set $\{*\}$ for Set, a trivial group for Grp and $AbGrp$, a singleton topological space for Top, the zero module 0 for $R\text{-}Mod$ (and similarly with Vec_k), and the zero ring for $Ring$.

Thus, e.g., for any topological space Y, giving a continuous map $f_i : Y \to X_i$ for each i is equivalent to giving a single continuous map $f : Y \to \prod_i X_i$, such that for each i, f projects along the i-th factor to f_i. This is why the product topology was defined the way

it was: all those basic open sets etc. By the uniqueness of products (Exercise 3.17), this was the only way the product topological space could have been defined.

For those of you who are familiar with algebraic varieties, the category of algebraic varieties over k has binary products, which is the 'usual' binary product of algebraic varieties.

Exercise 3.19. In the category of fields, products do not exist, nor does the category have a terminal object.

Exercise 3.20. Prove the following enhancement of the fact that products exist in Set: For any category $\mathcal{C}$, the category $\mathrm{Presh}(\mathcal{C}) = \mathrm{Fun}(\mathcal{C}^{op}, Set)$ has arbitrary small products.

More precisely, given functors $(F_i : \mathcal{C} \rightsquigarrow Set)_{i\in I}$, take $\prod_{i\in I} F_i$ to be the functor $F : \mathcal{C} \rightsquigarrow Set$ such that for each $X \in \mathrm{Ob}\,\mathcal{C}$,

$$F(X) = \prod_{i\in I} F_i(X), \qquad \text{this product being taken in } Set.$$

It is clear how to complete the definition of F by defining it for morphisms, and it is clear how to define the $\pi_i : F \to F_i$.

Remark 3.21. (i) Recall the Yoneda embedding $\mathcal{C} \rightsquigarrow Presh(\mathcal{C})$. What Definition 3.22 (specifically, (12)) tells us is that $(X, (\pi_i : X \to X_i))$ is a product of $(X_i)_{i\in I}$ if and only if $(- \circ \pi_i)_{i\in I}$ induces, for all $Y \in \mathrm{Ob}\,\mathcal{C}$:

$$h^X(Y) \to \prod_{i\in I} h^{X_i}(Y).$$

Upshot: $(X, (\pi_i : X \to X_i)_{i\in I})$ is a product of $(X_i)_{i\in I}$ if and only if $(h^X, (- \circ \pi_i)_{i\in I})$ is a product of $(h^{X_i})_{i\in I}$ in $\mathrm{Presh}(\mathcal{C})$.

(ii) This gives us a slightly different way to define a product of $(X_i)_{i\in I}$, as follows. Consider the functor $F := \prod_i h^{X_i}$ in $Presh(\mathcal{C})$, defined using the explicit set-theoretic description of Exercise 3.20, so it is given (at the level of objects) by

$$Y \mapsto \prod_{i\in I} h^{X_i}(Y).$$

If this functor is representable by some object X, then we say that a product of $(X_i)_{i\in I}$ exists, and define a product of $(X_i)_{i\in I}$ to be X together with a choice of a natural isomorphism

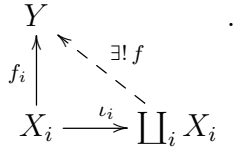
$$h^X \to \prod_{i\in I} h^{X_i}. \tag{13}$$

Why is this equivalent to the definition in Definition 3.22? Here is the rough idea. As mentioned in Remark 3.15, giving an equivalence of the form (13) is equivalent to giving the image u of $\mathrm{id}_X \in h^X(X)$ in $F(X) = \prod_{i\in I} h^{X_i}(X) = \prod_{i\in I} \mathrm{Mor}_{\mathcal{C}}(X, X_i)$. If this image is $u = (\pi_i)_{i\in I} \in \prod_{i\in I} \mathrm{Mor}_{\mathcal{C}}(X, X_i)$, then you can verify that the condition of Definition 3.22 is satisfied with $(X, (\pi_i)_{i\in I})$. Thus, to repeat, $(\pi_i)_{i\in I}$ is exactly the 'universal object' u of Remark 3.15.

Rough summary of this discussion: $\mathcal{C}$ may not be closed under small products, but $\mathcal{C}$ is contained in a category which is closed under small products, namely $\mathrm{Presh}(\mathcal{C})$. Thus, $\prod_{i\in I} h^{X_i}$ always exists, and its representability is equivalent to the existence of $\prod_{i\in I} X_i$. Another way to look at this is that in $\mathrm{Presh}(\mathcal{C})$, you can define products set-theoretically, and *define* products in $\mathcal{C}$ by 'restriction' from $\mathrm{Presh}(\mathcal{C})$ using the Yoneda embedding.

3.5. **Coproducts.** Here things are as in Subsection 3.4, but with all arrows reversed, so we will be relatively brief.

Definition 3.22. (i) A coproduct, or a categorical sum, of a family $(X_i)_{i\in I}$ in a category $\mathcal{C}$ is a pair $(X = \coprod_{i\in I} X_i, (\iota_i)_{i\in I})$, where $X \in \mathrm{Ob}\,\mathcal{C}$ and $\iota_i : X_i \to X$ in $\mathcal{C}$ for each i, such that given $Y \in \mathrm{Ob}\,\mathcal{C}$ and morphisms $f_i : X_i \to Y$ for each $i \in I$, there exists a unique morphism $f : X \to Y$ such that for each $i \in I$, the following diagram commutes:



In other words, for each $Y \in \mathrm{Ob}\,\mathcal{C}$ we have a bijection

$$\mathrm{Mor}_{\mathcal{C}}(\coprod_i X_i, Y) \overset{\iota_i \circ -}{\to} \prod_i \mathrm{Mor}_{\mathcal{C}}(X_i, Y). \tag{14}$$

When $I = \{1, 2\}$, we get the special case of a 'binary coproduct'.

(ii) We can interpret or extend the definition in (ii) to apply in the case where $I = \varnothing$: An 'empty coproduct' in a category $\mathcal{C}$ is by definition an *initial object* or a *coterminal object* in the category, by which one means an object $X \in \mathrm{Ob}\,\mathcal{C}$ such that for all $Y \in \mathrm{Ob}\,\mathcal{C}$, $\mathrm{Mor}_{\mathcal{C}}(X, Y)$ is a singleton.

(iii) We say that a category $\mathcal{C}$ has small coproducts if every collection $(X_i)_{i\in I}$ of objects of $\mathcal{C}$ has a coproduct (since the set I is allowed to be the empty set, this includes the requirement that $\mathcal{C}$ has an initial object).

A coproduct in a category need not exist, but if it does, it is suitably unique:

Exercise 3.23. Formulate and prove an analogue of Exercise 3.17 for coproducts.

Again, it is the uniqueness of the coproduct (Exercise 3.23), that justifies writing $\coprod_i X_i$ for the (object underlying the tuple constituting) the coproduct of the X_i (and similarly $X_1 \coprod X_2$ for a binary coproduct).

Example 3.24. The following categories have arbitrary small coproducts:

- In *Set*, coproduct is given by the disjoint union: $\coprod_i X_i$ can be taken to be the disjoint union X of the X_i, and $\iota_j : X_j \to X$ to be obvious inclusion. Of course, one also needs to remark that *Set* does have an initial object, which is $\varnothing$.

- In *Grp*, coproduct is given by the free product: the free product $G * H$ of G and H is the sequence of words $s_1 \cdots s_n$ with each s_i belonging to G or H, with obvious redundancies removed: e.g., "identity elements can be removed" and two successive terms belonging to the same group can be multiplied together. In other words, each of its elements can be identified with either the empty word, or an alternating sequence of elements of G and H. The maps $G \to G * H$ and $H \to G * H$ are obvious. *Grp* does have an initial object, the trivial group.
- For *Ring*, a coproduct exists, and is analogous to a free product sort of construction, but we will not bother with it (not least because I haven't looked it up myself). $\mathbb{Z}$ is an initial object in *Ring* (the 0 ring cannot be an initial object, because by definition, ring homomorphisms are required to send 1 to 1).
- In *AbGrp*, *R-Mod* and Vec_k, coproduct is given by direct sum (and the trivial group or the 0 group or module or vector space is the initial object). Thus, in all these categories, finite coproducts and finite products can be identified with each other, though not infinite ones.
- In *Top*, again, coproduct is given by the disjoint union, though make sure you know how to define the 'correct' topology on $\bigsqcup_i X_i$: the $X_i \subset X$ are all open and disjoint, and the topology of each X_i coincides with the one it gets from the inclusion $X_i \subset X$. Again, $\varnothing$ serves as an initial object.

Thus, e.g., for any topological space Y, giving a continuous map $f_i : X_i \to Y$ for each i is equivalent to giving a single continuous map $f : \coprod_i X_i \to Y$, such that for each i, f restricts to X_i as f_i.

Remark 3.25. A coproduct in $\mathcal{C}$ is essentially a product in $\mathcal{C}^{op}$, so we can use the Yoneda embedding $h_\bullet$ for $\mathcal{C}^{op}$, $h_\bullet : \mathcal{C}^{op} \rightsquigarrow \mathrm{Presh}(\mathcal{C}^{op}) = \mathrm{Fun}(\mathcal{C}, Set)$, to get a description of coproduct in $\mathcal{C}$. Namely, Definition 3.22 (specifically, (14)) tells us that $(X, (\iota_i : X_i \to X))$ is a coproduct of $(X_i)_{i \in I}$ if and only if $(- \circ \iota_i)_{i \in I}$ induces, for all $Y \in \mathrm{Ob}\,\mathcal{C}$:

$$h_X(Y) \to \prod_{i \in I} h_{X_i}(Y).$$

Upshot: $(X, (\iota_i : X_i \to X))$ is a coproduct of $(X_i)_{i \in I}$ if and only if $(h_X, (- \circ \iota_i)_{i \in I})$ is a product of $(h_{X_i})_{i \in I}$ in $\mathrm{Fun}(\mathcal{C}, Set) = \mathrm{Presh}(\mathcal{C}^{op})$. Appropriate analogues of the rest of Remark 3.21 also apply here, e.g., the existence of $\coprod_i X_i$ is equivalent to the representability of $\prod_i h_{X_i}$.

However, note some asymmetry between the two situations, which basically arises from covariance vs contravariance, and is also reflected in the difference between (12) and (14): there is a '$\prod$' on both sides of the former, but the latter has a '$\coprod$' on the left-hand side and a '$\prod$' on the right-hand side.

4. Lecture 4 – Limits and colimits

We continue with the convention that, unless otherwise stated, any category that we will encounter is locally small, though we will make an exception for presheaf categories of categories we work with.

4.1. **Monomorphisms and epimorphisms.**

Definition 4.1. (i) A morphism $f : X \to Y$ in a category $\mathcal{C}$ is said to be a monomorphism if it has "left-cancellation", i.e., if $g_1, g_2 : Z \to X$ are such that $f \circ g_1 = f \circ g_2 : Z \to Y$, then $g_1 = g_2$. (Equivalently: $h^X \to h^Y$ is objectwise injective).

(ii) A morphism $f : X \to Y$ is said to be an epimorphism if it has "right-cancellation", i.e., if $g_1, g_2 : Y \to Z$ are such that $g_1 \circ f = g_2 \circ f : X \to Z$, then $g_1 = g_2$.

Thus, $f : X \to Y$ in $\mathcal{C}$ is a monomorphism if and only if, viewed as a morphism in $\mathcal{C}^{op}$, it is an epimorphism.

Example 4.2. (i) In Set, a morphism $f : X \to Y$ is a monomorphism (resp., epimorphism) if and only if it is an injective (resp., surjective) function.

(ii) The "if" part of the analogous assertion is true in $Grp, AbGrp, Ring, R\text{-}Mod, Vec_k$ and Top, and also in the full subcategory $HausTop$ of Top consisting of the Hausdorff topological spaces. This can be viewed more category-theoretically: if $F : \mathcal{C} \rightsquigarrow Set$ is a faithful functor, then $f : X \to Y$ is a monomorphism (resp., epimorphism) whenever the map $F(f) : F(X) \to F(Y)$ of sets is (for these categories, this is true with F the forgetful functor to Set).

(iii) However, the "only if" part, while true for $Grp, Top, AbGrp, R\text{-}Mod$ and Vec_k (a bit of work is needed to show this for Grp and Top), is not true for $Ring$ or $HausTop$: it is an easy exercise to check that every monomorphism is injective in these categories (as also in Grp, Top), but epimorphisms may not be surjective in $Ring$ or $HausTop$: In $Ring$, $\mathbb{Z} \to \mathbb{Q}$ is an epimorphism, [5] while in $HausTop$, any morphism with a dense image is an epimorphism (easy but good exercise).

4.2. **Equalizers and coequalizers.**

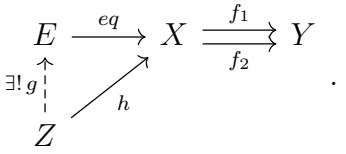
An 'equalizer' of $f_1, f_2 : X \to Y$ tries to capture the notion of the 'subset of X where f_2 and f_2 agree'. Formally:

Definition 4.3. Let $f_1, f_2 : X \to Y$ be morphisms in $\mathcal{C}$.

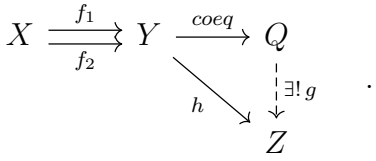
(i) An equalizer of f_1 and f_2 is a morphism $eq : E \to X$ in $\mathcal{C}$, satisfying $f_1 \circ eq = f_2 \circ eq$, and satisfying the following universal property: for any morphism $h : Z \to X$ such

[5] While dealing with the category of rings, please keep in mind that ring homomorphisms are required to send 1 to 1.

that $f_1 \circ h = f_2 \circ h$, there exists a unique morphism $g : Z \to E$ such that $h = eq \circ g$:



(ii) A coequalizer of f_1 and f_2 is a morphism $coeq : Y \to Q$, satisfying $coeq \circ f_1 = coeq \circ f_2$, and satisfying the following universal property: for any morphism $h : Y \to Z$ such that $h \circ f_1 = h \circ f_2$, there exists a unique morphism $g : Q \to Z$ such that $h = g \circ coeq$:



Exercise 4.4. Formulate a statement that captures "Equalizers and coequalizers are uniquely unique", and prove it.

Please note that the equalizer and the coequalizer refer to the morphisms $eq : E \to X$ and $coeq : Y \to Q$, and not just to the objects E and Q. However, by abuse of notation, we might often write just E and Q when the maps $eq : E \to X$ and $coeq : Y \to Q$ are clear.

Example 4.5. (i) In Set, an equalizer E of $f_1, f_2 : X \to Y$ is simply the (largest) subset of X where f_1 and f_2 agree, while their coequalizer Q is the quotient of Y by the equivalence relation generated by the relation "$f_1(x) \sim f_2(x)$ for all $x \in X$". In Top, the prescriptions for the equalizer E and the coequalizer Q are the same as for Set, except that E should be given the induced topology from X and that Q should be given the quotient topology dictated by $Y \to Q$.

(ii) In Grp, the equalizer is as in Set (but it forms a subgroup of X and one remembers it as a subgroup, not a subset), and the coequalizer of $f_1, f_2 : X \to Y$ is the quotient of Y by the smallest *normal* subgroup of Y containing $f_1(x)f_2(x)^{-1}$ for each $x \in X$.

(iii) In $AbGrp, R\text{-}Mod$ and Vec_k, the equalizer of f_1 and f_2 is the kernel of $f_1 - f_2$, and their coequalizer is the cokernel of $f_1 - f_2$ (thus, for $AbGrp$, the equalizer and the coequalizer are the same as in Grp, only the description simplifies due to abelianness).

(iv) In Presh($\mathcal{C}$), the equalizer is the "object-wise set-theoretic equalizer" (make this precise and prove it).

Exercise 4.6. (i) Show that every equalizer is a monomorphism, and that every coequalizer is an epimorphism. The converse does not hold, but let us not bother about that now.

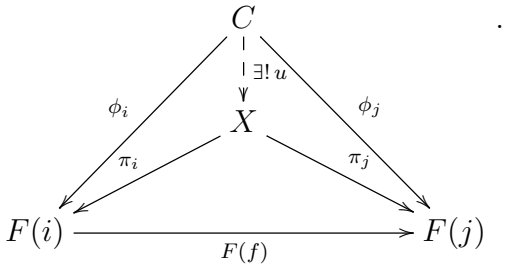
(ii) Describe equalizers using the Yoneda embedding and the obvious description of equalizers in Presh($\mathcal{C}$) = Fun($\mathcal{C}^{op}, Set$), in the style of Remark 3.21. Similarly, describe coequalizers as in Remark 3.25: note that this involves an equalizer rather than a coequalizer in the presheaf category.

(iii) Example from wikipedia: Show that in Top, the coequalizer of the two maps $f_1, f_2 : \{*\} \to [-1, 1]$, where $f_1(*) = 0$ and $f_2(*) = 1$, is S^1.

4.3. **Limits and colimits.** Notice that there were some parallels between products and equalizers; limits are a simultaneous generalization of these concepts. Similarly, colimits simultaneously generalize coproducts and coequalizers.

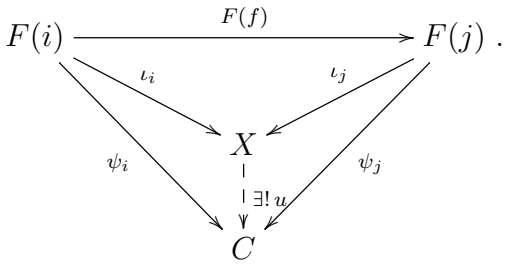
Notation 4.7. Note that given any functor $F : J \rightsquigarrow \mathcal{C}$, it can be also viewed as a functor $F^{op} : J^{op} \rightsquigarrow \mathcal{C}^{op}$: this defines F^{op}.

Definition 4.8. (i) (Cones) Let $F : J \rightsquigarrow \mathcal{C}$ be a functor. A cone to F is a pair $(C, (\phi_j)_{j \in \mathrm{Ob}\, J})$ consisting of an object $C \in \mathrm{Ob}\, \mathcal{C}$, together with morphisms $\phi_j : C \to F(j)$, for all $j \in J$, such that for all $f : i \to j$ in J, we have $F(f) \circ \phi_i = \phi_j$; see the outer triangle of the following diagram.



(ii) (Limits) A limit F is then a 'universal cone to F', i.e., it is a cone $(X, (\pi_j)_{j \in \mathrm{Ob}\, J})$ to F, such that given any cone $(C, (\phi_j)_{j \in \mathrm{Ob}\, J})$ to F, there exists a unique morphism $u : C \to X$ such that for all $j \in J$, we have $\pi_j = u \circ \phi_j$ (see the above diagram).

(iii) (Cocones) Let $F : J \rightsquigarrow \mathcal{C}$ be a functor. A cocone to F is a cone to $F^{op} : J^{op} \rightsquigarrow \mathcal{C}^{op}$. In other words, it is a pair $(C, (\psi_j)_{j \in \mathrm{Ob}\, J})$ consisting of an object $C \in \mathrm{Ob}\, \mathcal{C}$, together with morphisms $\psi_j : F(j) \to C$, for all $j \in J$, such that for all $f : i \to j$ in J, we have $\psi_j \circ F(f) = \psi_i$; see the outer triangle of the following diagram.



(iv) (Colimits) A colimit of F is a limit of $F^{op} : J^{op} \to \mathcal{C}^{op}$.
In other words, it is a 'universal cocone to F', i.e., a cocone $(X, (\iota_j)_{j \in \mathrm{Ob}\, J})$, such that given any cocone $(C, (\psi_j)_{j \in \mathrm{Ob}\, J})$, there exists a unique morphism $u : X \to C$ such that for all $j \in J$, we have $u \circ \iota_j = \psi_j$ (see the above diagram).

(v) By a small limit we refer to a limit of a functor $F : J \rightsquigarrow \mathcal{C}$, where J is a small category. Similarly, we define small colimits.

(vi) A category $\mathcal{C}$ is said to be complete if it is "closed under small limits" – i.e., if every functor $F : J \rightsquigarrow \mathcal{C}$, with J a small category, has a limit. Similarly, we call $\mathcal{C}$ and cocomplete if it is closed under small colimits.

Remark 4.9. If J is the empty category (no object, no morphism), then there is a unique functor $F : J \rightsquigarrow \mathcal{C}$. In this case, a cone to F is just an object of $\mathcal{C}$, so that a limit of F is just a final object of $\mathcal{C}$ (this sentence means: neither a limit of F nor a final object of $\mathcal{C}$ may exist, but one exists if and only if the other does, and when they do they are uniquely isomorphic). Similarly, a colimit of F is just an initial object of $\mathcal{C}$.

Exercise 4.10. (i) Let $F : J \rightsquigarrow \mathcal{C}$ be a functor. Organize the cones to F into a category by defining morphisms between them (this is vaguely formulated, but I expect you to formulate the most natural notion of morphisms between these). Show that a limit of F is the same as a final object of this category. This gives another way to phrase the definition of a limit of F.

(ii) Do (i) with cones and limit replaced by cocones and colimit.

(iii) Using (i) and (ii) or otherwise, rigorously formulate and prove the following assertion: a limit of F is uniquely unique if it exists, as is a colimit of F.

(iv) (Limits generalize products) Let $(X_j)_{j\in J}$ be a set of objects of $\mathcal{C}$ indexed by J. We view J as a discrete category, which means: its collection of objects is the set J, and its only morphisms are the identity morphisms (the "discreteness"), so that there is a unique functor $F : J \rightsquigarrow \mathcal{C}$ sending each j to X_j. Show that a limit of F is the same as a product of $(X_j)_{j\in \operatorname{Ob} J}$. This applies even when J is empty (see Remark 4.9 above, which is therefore a special case of this exercise).

(v) (Limits generalize equalizers) Consider the situation of Definition 4.3, so we have $f_1, f_2 : X \to Y$ in $\mathcal{C}$. Let J be a category consisting of two objects a, b, and which has exactly two non-identity morphisms, denoted $\tilde{f}_1, \tilde{f}_2 : a \to b$. Define $F : J \to \mathcal{C}$ so that $F(a) = X, F(b) = Y, F(\tilde{f}_1) = f_1$ and $F(\tilde{f}_2) = f_2$ (and it takes identity morphisms to the appropriate identity morphisms) (see the diagram below). Show that a limit of F is the same as an equalizer of f_1 and f_2.

$$a \underset{\tilde{f}_2}{\overset{\tilde{f}_1}{\rightrightarrows}} b \qquad \overset{F}{\rightsquigarrow} \qquad X \underset{f_2}{\overset{f_1}{\rightrightarrows}} Y$$

(vi) (Colimits generalize coproducts and coequalizers) Do the colimit analogue of (iv) and (v).

Example 4.11. (i) In *Set*, all small limits exist. Namely, if $F : J \rightsquigarrow \mathcal{C}$ is a functor with J a small category, show as an exercise that a limit of F is given by $(X, (\pi_j)_{j\in \operatorname{Ob} J})$, where

$$X = \Big\{(x_j)_j \in \prod_{j\in \operatorname{Ob} J} F(j) \mid \forall\, f : i \to j \text{ in } J, F(f)(x_i) = x_j\Big\}, \tag{15}$$

and $\pi_j : X \to F(j)$ is given by projection to the j-th factor in $\prod_{j\in \operatorname{Ob} J} F(j)$. Thus, *Set* is complete. As this example illustrates, a limit is formed by products and

equalizers – as we now formalize in the following form that will be useful later: a limit of F is given by the equalizer of the following diagram:

$$\prod_{j \in \mathrm{Ob}\, J} F(j) \underset{g_2}{\overset{g_1}{\rightrightarrows}} \prod_{f \in \mathrm{Mor}(J)} F(t(f)), \tag{16}$$

where:

- $\mathrm{Mor}(J)$ is the set of all morphisms in J;
- for $f : i \to j$ in J, we have written $s(f) = i$ and $t(f) = j$ (s for "source" and t for "target");
- $g_1((x_j)_{j \in \mathrm{Ob}\, J}) = (x_{t(f)})_{f \in \mathrm{Mor}(J)}$ and $g_2((x_j)_{j \in \mathrm{Ob}\, J}) = (F(f)(x_{s(f)}))_{f \in \mathrm{Mor}(J)}$ (thus, in the condition $F(f)(x_i) = x_j$ in (15), $x_{t(f)}$ captures x_j, and $F(f)(x_{s(f)})$ captures $F(f)(x_i)$).

(ii) A similar prescription works for $Grp, AbGrp, R\text{-}Mod, Vec_k, Ring$ and Top, with the difference that one has to keep the extra structures in mind; e.g., for everything other than Top one uses component-wise operations, and for Top gives X the induced topology from the product topology: in other words, the weakest topology on X such that each $\pi_j : X \to X_j$ is continuous. Thus, $Grp, AbGrp, R\text{-}Mod, Vec_k, Ring$ and Top are complete.

(iii) In Set, all small colimits exist as well. Namely, if $F : J \to Set$ is a functor, with J a small category, one shows that a colimit of F is given by $(X, (\iota_j)_{j \in J})$, where:

$$X = \left(\bigsqcup_{j \in \mathrm{Ob}\, J} F(j) \right) / \sim,$$

where the equivalence relation $\sim$ is generated by the requirement that $x_i \sim x_j$ whenever $x_i \in F(i), x_j \in F(j)$, and there exists $f : i \to j$ such that $x_j = F(f)(x_i)$. As with limits above, this colimit is seen to be built out of coproducts and coequalizers. Thus, Set is both complete and cocomplete.

(iv) A similar prescription gives small colimits in Top (involving the use of quotient topology), so that Top is both complete and cocomplete.

(v) It turns out that Grp and $Ring$ are complete and cocomplete too, but the constructions for Set and Top do not work for these. I have never bothered looking up or working it out.

(vi) However, in $AbGrp, R\text{-}Mod$ and Vec_k, small colimits are easy to define, by appropriately adapting colimits for Set: replace $\bigsqcup$ by $\bigoplus$, and 'equivalence relation generated by' by 'subgroup/submodule/subvector space generated by' (in other words, "Span of"). So these categories are complete and cocomplete as well.

(vii) For any category $\mathcal{C}$, $\mathrm{Presh}(\mathcal{C})$ is both complete and cocomplete: one constructs limits and colimits "object-wise", using the limits and colimits for Set. In particular, one can construct a limit exactly as in (16), using products and an equalizer, namely:

$$\prod_{j \in \mathrm{Ob}\, J} F(j) \underset{g_2}{\overset{g_1}{\rightrightarrows}} \prod_{f \in \mathrm{Mor}(J)} F(t(f)), \tag{17}$$

with notation as in (16).

Exercise 4.12. Show that Remarks 3.21 and 3.25, and Exercise 4.6(ii), generalize to describe limits and colimits. Namely:

(i) The Yoneda embedding $h^\bullet : \mathcal{C} \rightsquigarrow \mathrm{Presh}(\mathcal{C})$ preserves limits. In fact, a functor $F : J \rightsquigarrow \mathcal{C}$ has a limit in $\mathcal{C}$ if and only if the limit of $h^\bullet \circ F : J \rightsquigarrow \mathrm{Presh}(\mathcal{C})$ (which exists since $\mathrm{Presh}(\mathcal{C})$ has been observed above in Example 4.11(vii) to be complete) is representable. Thus, as in Remark 3.21, this can give another definition of a limit.

(ii) Since the Yoneda embedding $h_\bullet : \mathcal{C}^{op} \rightsquigarrow \mathrm{Presh}(\mathcal{C}^{op}) = \mathrm{Fun}(\mathcal{C}, Set)$ preserves limits, it takes colimits of $\mathcal{C}$ (which are limits in $\mathcal{C}^{op}$) to limits in $\mathrm{Fun}(\mathcal{C}, Set)$.

A special case of the above exercise, which is much more simple and immediate, and yet an important result, is:

Theorem 4.13. *If $X \in \mathrm{Ob}\,\mathcal{C}$, then the functor $\mathrm{Mor}_{\mathcal{C}}(X, -) : \mathcal{C} \rightsquigarrow Set$ as well as the functor $\mathrm{Mor}_{\mathcal{C}}(-, X) : \mathcal{C}^{op} \rightsquigarrow Set$ preserve small limits.*

Proof. Easy exercise; this is pretty much the definition of a limit and a colimit. Here, we say that a functor $G : \mathcal{C} \rightsquigarrow \mathcal{D}$ 'preserves small limits' if for any functor $F : J \rightsquigarrow \mathcal{C}$ with J a small category, if $(X, (\pi_j)_j)$ is a limit of F, then $(G(X), (G(\pi_j))_j)$ is a limit of the composite functor $G \circ F : J \rightsquigarrow \mathcal{D}$. □

Remark 4.14. Later we will hopefully see that Theorem 4.13 contains the left-exactness of the 'Hom-functors' in the context of R-modules.

4.4. **Direct and inverse limits.** We now discuss how "direct limits" and "inverse limits" are special cases of the above constructions. Let $(J, \leqslant)$ be a directed set: J is a nonempty set, and "$\leqslant$" is a reflexive and transitive relation on it such that any two elements of J have an upper bound. Antisymmetry is not always imposed, but let us impose it because I don't want to worry about the nuances. One makes $(J, \leqslant)$ into a category: its objects are just the elements of J, and for $i \neq j$ in J, there is a single morphism $i \to j$ if $i \leqslant j$, and none otherwise. Call this category J as well. Then we can consider limits of functors $G : J^{op} \to \mathcal{C}$ and colimits of functors $F : J \rightsquigarrow \mathcal{C}$.

Limits of such functors $G : J^{op} \rightsquigarrow \mathcal{C}$ are called inverse limits, denoted $\varprojlim_j G(j)$, and colimits of such functors $F : J \rightsquigarrow \mathcal{C}$ are called direct limits or directed colimits, denoted $\varinjlim_j F(j)$. We will prefer saying 'directed colimits' for the latter, since those are colimits in the sense defined above.

Let us be slightly more explicit; show as an easy exercise the following remark:

Remark 4.15. (i) Giving such a functor $F : J \rightsquigarrow \mathcal{C}$ as above is equivalent to giving a family $(X_j)_{j \in J}$ of objects of $\mathcal{C}$ together with a family of morphisms $\{\varphi_{ji} : X_i \to X_j\}_{i,j \in J, i \leqslant j}$ such that $\varphi_{jj} = \mathrm{id}\ \forall\, j \in J$, and $\varphi_{ki} = \varphi_{kj} \circ \varphi_{ji}$ whenever $i \leqslant j \leqslant k$ in J.

(ii) Giving a functor $G : J^{op} \rightsquigarrow \mathcal{C}$ as above is equivalent to giving a family $(X_j)_{j\in J}$ of objects of $\mathcal{C}$ together with a family of morphisms $\{\psi_{ij} : X_j \to X_i\}_{i,j\in J, i\leqslant j}\}$ such that $\psi_{jj} = \mathrm{id}\ \forall\, j \in J$, and $\psi_{ij} \circ \psi_{jk} = \psi_{ik}$ whenever $i \leqslant j \leqslant k$ in J.

This motivates:

Definition 4.16. (i) A direct system over $(J, \leqslant)$ is a pair $((X_j)_{j\in J}, \{\varphi_{ji} : X_i \to X_j\}_{i,j\in J, i\leqslant J})$ as in Remark 4.15(i) above. Given such a direct system $((X_j)_j, (\varphi_{ji})_{i\leqslant j})$, the colimit of the functor $F : J \rightsquigarrow \mathcal{C}$ associated to it as above is called a directed colimit of this directed system, and typically written

$$\varinjlim_j X_j$$

(by convention we have suppressed the maps φ_{ji} from the notation, though this limit depends on which φ_{ji} we are using).

(ii) An inverse system over $(J, \leqslant)$ is a pair $((X_j)_{j\in J}, \{\psi_{ij} : X_j \to X_i\}_{i,j\in J, i\leqslant j})$ as in Remark 4.15(ii) above. Given such an inverse system $((X_j)_j, (\psi_{ij})_{i\leqslant j})$, the colimit of the functor $G : J^{op} \rightsquigarrow \mathcal{C}$ associated to it as above is called an inverse limit of this inverse system, and typically written

$$\varprojlim_j X_j.$$

In the context of direct and inverse systems, we can translate the descriptions of limits and colimits from Example 4.11 to the language above:

- Let $((X_j)_j, (\varphi_{ji})_{i\leqslant j})$ be a direct system over $(J, \leqslant)$. If additionally $\mathcal{C}$ is *AbGrp*, *R-Mod* or Vec_k, we can write (the object underlying) $\varinjlim_j X_j$ as:

$$\left(\bigoplus_{j\in J} X_j\right) / \mathrm{Span}(\{x_j - \varphi_{ji}(x_i) \mid i, j \in J, i \leqslant j\}). \tag{18}$$

  For *Set* or *Top*, replace the direct sum $\bigoplus$ by the disjoint union $\bigsqcup$, and *Span* by "the equivalence relation generated by". Something like this can be done for *Grp* and *Ring*, but we will not get into it.
- Let $((X_j)_j, (\psi_{ij})_{i\leqslant j})$ be an inverse system over $(J, \leqslant)$. If additionally $\mathcal{C}$ is *Set*, *Top*, *Grp*, *Ring*, *AbGrp*, *R-Mod* or Vec_k, we can write (the object underlying) $\varprojlim_j X_j$ as:

$$\left\{(x_j) \in \prod_{j\in J} X_j \mid x_i = \psi_{ij}(x_j)\ \forall\, i, j \in J, i \leqslant j\right\}. \tag{19}$$

Example 4.17. (i) Let $((X_j)_j, (\varphi_{ji})_{i\leqslant j})$ be a direct system over $(J, \leqslant)$. Assume that $\mathcal{C}$ is $Set, Top, Grp, AbGrp, R\text{-}Mod, Vec_k, Ring$, or the category of fields. In these settings, we can make sense of containments such as $X \subset Y$. Assume that there is some object X in $\mathcal{C}$ such that for all $i \leqslant j$ in J, we have $X_i \subset X_j \subset X$, and

$\varphi_{ji} : X_i \to X_j$ is the inclusion. Then it follows from (18) that the object underlying $\varinjlim_j X_j$ identifies with

$$\bigcup_{j \in J} X_j,$$

this union taken inside X (and the morphisms $\iota_j : X_j \to X$ are the obvious containments too). Thus, a directed colimit is a generalization of union.

(ii) Let $((X_j)_j, (\psi_{ij})_{i \leqslant j})$ be an inverse system over $(J, \leqslant)$. Assume that $\mathcal{C}$ is $\mathcal{C}$ is Set, Top, Grp, $AbGrp$, $R\text{-}Mod$, Vec_k, $Ring$, or the category of fields. Assume that there is some object X in $\mathcal{C}$ such that for all $i \leqslant j$ in J, we have $X_j \subset X_i \subset X$, and $\psi_{ij} : X_j \to X_i$ is the inclusion. Then it follows from (19) that the object underlying $\varprojlim_j X_j$ identifies with

$$\bigcap_{j \in J} X_j,$$

so an inverse limit can be thought of as a generalization of intersection.

Example 4.18. We now discuss some examples of directed colimits.

(i) The category of fields is bad in so many ways (e.g., it does not have binary products or coproducts), but it does have directed colimits. If L/K is an infinite algebraic extension, one can take J to be the set of finite extensions of K contained in L, ordered under inclusion, and $X_j := j$ for all $j \in J$, and $\varphi_{ji} : X_i \to X_j$ to be the inclusion $i \hookrightarrow j$ (which makes sense as $i \leqslant j$). This is a special case of Example 4.17(i), and gives the directed colimit of the X_j as L. Thus, L is a directed colimit of finite extensions of K.

(ii) We take $(J, \leqslant)$ to be $\mathbb{N}_{\geqslant 1}$ with the partial order where $m \leqslant n$ if and only if $m|n$. Set $X_n = (1/n)\mathbb{Z}$ for each n, and for $m \leqslant n$, let $\varphi_{nm} : X_m \to X_n$ be the inclusion. Then it follows from Example 4.17(i) that $\varinjlim_n X_n \cong \mathbb{Q}$ (by abuse of notation: to complete it one should say the X_n map to this by inclusion).

(iii) We take $(J, \leqslant)$ to be as in (ii), set $Y_n = \mathbb{Z}$ for each n, and for $m \leqslant n$, we set $\varphi'_{nm} : Y_m = \mathbb{Z} \to \mathbb{Z} = Y_n$ to be multiplication by $n/m \in \mathbb{Z}$. To compute $\varinjlim_n Y_n$, letting the X_n and the φ_{nm} be as in (ii), one considers the isomorphisms $f_n : Y_n \to X_n$ given by $x \mapsto x/n$, and then notice that $\varphi_{nm} \circ f_m = f_n \circ \varphi'_{nm}$. This allows us to identify the directed system $((Y_n)_n, (\varphi'_{nm})_{m \leqslant n})$ with the directed system $((X_n)_n, (\varphi_{nm})_{m \leqslant n})$ in (ii) above, and we get

$$\varinjlim_n Y_n \cong \varinjlim_n X_n \cong \mathbb{Q}.$$

(iv) We also have the following variants of the examples in (ii) and (iii): take the same $(J, \leqslant)$, set $X_n = (\frac{1}{n}\mathbb{Z}/\mathbb{Z})$ and $Y_n = \mathbb{Z}/n\mathbb{Z}$ for each n. One can define the φ_{nm} and the φ'_{nm} exactly as in (ii) and (iii). Then the same arguments as in these examples

give:

$$\varinjlim_n \mathbb{Z}/n\mathbb{Z} \cong \varinjlim_n (\frac{1}{n}\mathbb{Z})/\mathbb{Z} \cong \mathbb{Q}/\mathbb{Z}.$$

Please note that this is nothing peculiar: this is just expressing $\mathbb{Q}/\mathbb{Z}$ as a directed colimit/union of cyclic groups in a straightforward way.

Example 4.19. We now discuss some examples of inverse limits.

(i) (p-adic integers) Take $(J, \leqslant)$ to be $\mathbb{N}_{\geqslant 1}$ together with the usual order, and $\mathcal{C}$ to be *Ring* (one could also view this example with *AbGrp* instead). Fix a prime number p. Take $X_n = \mathbb{Z}/p^n\mathbb{Z}$, and for $m \leqslant n$, let $\psi_{mn} : \mathbb{Z}/p^n\mathbb{Z} \to \mathbb{Z}/p^m\mathbb{Z}$ be the obvious map. Then, by (19), the inverse limit of this system is the ring

$$\mathbb{Z}_p := \varprojlim_n \mathbb{Z}/p^n\mathbb{Z} = \Big\{(a_n)_n \in \prod_n \mathbb{Z}/p^n\mathbb{Z} \mid a_n \in \mathbb{Z}/p^n\mathbb{Z} \text{ reduces to } a_m \in \mathbb{Z}/p^m\mathbb{Z} \text{ whenever } m \leqslant n\Big\}$$

(with component-wise addition and multiplication). It is an easy exercise to see that this ring $\mathbb{Z}_p$ is an integral domain. It is called the ring of p-adic integers, and its quotient field $\mathbb{Q}_p$ is called the field of p-adic numbers.

(ii) (Completion of a ring with respect to an ideal) Things are as in (i), but one replaces $\mathbb{Z}$ by a ring R and (p) by an ideal I, so sets $X_n = R/I^n$, and for $m \leqslant n$ defines $\psi_{mn} : R/I^n \to R/I^m$ to be the obvious map $R/I^n \to R/I^m$. Then, by (19), the inverse limit of this system is the ring

$$\hat{R} = \varprojlim_n R/I^n = \Big\{(a_n)_n \in \prod_n R/I^n \mid a_n \in R/I^n \text{ reduces to } a_m \in R/I^m \text{ whenever } m \leqslant n\Big\},$$

called *the completion of R with respect to I* ($\hat{R}$ depends on I, but one abuses notation by suppressing I from it).

(iii) (Formal power series ring) This is a special case of (ii) above. Check as an easy exericse that, taking $R[x]$ and (x) in place of R and (I), the completion we get is

$$\varprojlim_n R[x]/(x^n) = R[\![x]\!] := \Big\{\sum_{n=0}^{\infty} a_n x^n \mid a_n \in R \,\forall\, n\Big\},$$

the *formal power series ring in one variable over R*, with an obvious 'formal' addition and multiplication. Indeed, to see the above identification, note that an element of this inverse limit looks like (up to taking representatives in $R[x]$ for the $R[x]/(x^n)$) $(a_0, a_0 + a_1x, a_0 + a_1x + a_2x^2, \dots)$, so this sequence can be mapped to $\sum a_n x^n \in R[\![x]\!]$.

(iv) (The profinite completion of $\mathbb{Z}$) One again takes $J = \mathbb{N}_{\geqslant 1}$, but defines $m \leqslant n$ if $m|n$. One takes $X_n = \mathbb{Z}/n\mathbb{Z}$, and for $m \leqslant n$ one defines $\psi_{mn} : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ to be the obvious map. Then we get the ring

$$\varprojlim_n \mathbb{Z}/n\mathbb{Z} = \Big\{(a_n)_n \in \prod_n \mathbb{Z}/n\mathbb{Z} \mid a_n \in \mathbb{Z}/n\mathbb{Z} \text{ reduces to } a_m \in \mathbb{Z}/m\mathbb{Z} \text{ whenever } m|n\Big\},$$

usually denoted by $\hat{\mathbb{Z}}$, called in a group theoretic context the *profinite completion* of $\mathbb{Z}$.

(v) (Profinite completion) In a group theoretic context, one can similarly consider the profinite completion of other groups: J is to be taken as the directed set of finite index normal subgroups of the given group, ordered under reverse inclusion, and the inverse limit is usually given an appropriate topology, called the profinite topology, and viewed for practical reasons as a topological group.

The following exercise may be given as part of HW 2, though I am not sure yet.

Exercise 4.20. Use the Chinese remainder theorem to give an isomorphism

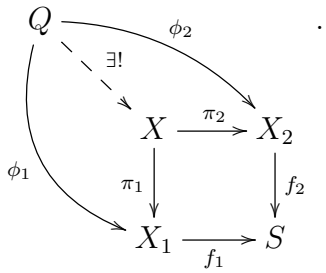
$$\hat{\mathbb{Z}} \cong \prod_p \mathbb{Z}_p.$$

4.5. **Pullbacks and pushouts.**

Definition 4.21. (i) (Pullback/fiber product/cartesian square) Given two morphisms $f : X_1 \to S$ and $f_2 : X_2 \to S$ in $\mathcal{C}$, a pullback or a fiber product (or fibered product) of f_1 and f_2 is the limit of the diagram $X_1 \xrightarrow{f_1} S \xleftarrow{f_2} X_2$, by which we mean the limit of a functor $F : J \to \mathcal{C}$, where J has three objects x_1, s and x_2, and exactly two nonidentity morphisms $\tilde{f_1} : x_1 \to s$ and $\tilde{f_2} : x_2 \to s$, and F sends $x_1, s, x_2, \tilde{f_1}, \tilde{f_2}$ respectively to X_1, S, X_2, f_1, f_2:

$$x_1 \xrightarrow{\tilde{f_1}} s \xleftarrow{\tilde{f_2}} x_2 \qquad \overset{F}{\rightsquigarrow} \qquad X_1 \xrightarrow{f_1} S \xleftarrow{f_2} X_2.$$

The pullback and its universal property are illustrated in the following diagram, where the triple (X, π_1, π_2) is the pullback:



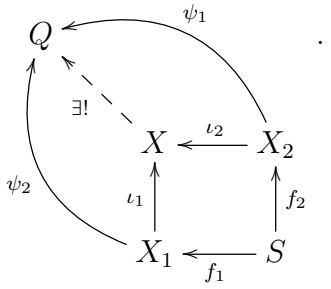
A diagram of the form of the square in the above diagram, where the morphisms from the top left object to the top right and the bottom left objects form a fiber product of the remaining two morphisms, is called a cartesian diagram.

(ii) (Pushout/fibered coproduct/cocartesian square) Given two morphisms $f_1 : S \to X_1$ and $f_2 : S \to X_2$ in $\mathcal{C}$, a pushout or a fibered coproduct of f_1 and f_2 is the colimit of the diagram $X_1 \xleftarrow{f_1} S \xrightarrow{f_2} X_2$, by which we mean the colimit of a functor $F : J \to \mathcal{C}$, where J has three objects x_1, s and x_2, and exactly two nonidentity

morphisms $\tilde{f}_1 : s \to x_1$ and $\tilde{f}_2 : s \to x_2$, and F sends $x_1, s, x_2, \tilde{f}_1, \tilde{f}_2$ respectively to X_1, S, X_2, f_1, f_2:

$$x_1 \xleftarrow{\tilde{f}_1} s \xrightarrow{\tilde{f}_2} x_2 \qquad \overset{F}{\rightsquigarrow} \qquad X_1 \xleftarrow{f_1} S \xrightarrow{f_2} X_2 \ .$$

The pushout and its universal property are illustrated in the following diagram, where the triple (X, ι_1, ι_2) is the pushout:



A diagram of the form of the square in the above diagram, where the morphisms to the top left object from the top right and the bottom left objects form a push out of the remaining two morphisms, is called a cocartesian diagram.

Remark 4.22. You see fibered coproducts in the Seifert-van Kampen theorem in topology. Fiber products are quite commonly seen in algebraic geometry. Some examples are special cases of what you have already seen for limits and colimits. For instance, in *Set* as in *Top*, given $f_1 : X_1 \to S$ and $f_2 : X_2 \to S$,

$$X_1 \times_S X_2 = \{(x_1, x_2) \mid x_1 \in X_1, x_2 \in X_2, \text{ and } f_1(x_1) = f_2(x_2) \in S\},$$

except that in the case of *Top* one should also give it the induced topology from the product topology.

4.6. **Criteria for existence of limits and colimits.** By a finite limit (resp., colimit), we mean the limit (resp., colimit) of a functor $F : J \rightsquigarrow \mathcal{C}$, where J is a finite category – i.e., the collection of objects of J is a finite set, and the collection of morphisms of J is also a finite set. A category which has all finite limits (resp., all finite colimits) is called finitely complete (resp., finitely cocomplete).

Theorem 4.23. *(i) A category is complete if and only if it has small products and equalizers.*

(ii) A category is cocomplete if and only if it has small coproducts and coequalizers.

Moreover, the above assertions are true with 'complete', 'small products', 'cocomplete' and 'small coproducts' replaced respectively by 'finitely complete', 'finite products', 'finitely cocomplete' and 'finite coproducts'.

Proof. We will prove (i) – then (ii) will follow from applying (i) to $\mathcal{C}^{op}$, and the proof of the 'finite' versions will be analogous.

"$\Rightarrow$" is immediate since small products and equalizers are special cases of small limits. Thus, let us prove "$\Leftarrow$", so assume that $\mathcal{C}$ has small products and equalizers. Let $F : J \rightsquigarrow \mathcal{C}$ be a functor, with J a small category. It is enough to show that F has a limit.

Motivated by (16), using analogous notation, we consider the equalizer of the following:

$$\prod_{j \in \mathrm{Ob}\, J} F(j) \underset{g_2}{\overset{g_1}{\rightrightarrows}} \prod_{f \in \mathrm{Mor}(J)} F(t(f)), \tag{20}$$

where g_1 is the unique map from the left-hand side to the right-hand side whose composition with the projection onto the f-th component, for $f \in \mathrm{Mor}(J)$, is the projection $\pi_{t(f)}$ from the left-hand side to its $t(f)$-factor (recall that these projections π_j come with the definition of a product), and g_2 is the map whose projection is $F(f) \circ \pi_{s(f)}$.

To show that this equalizer is a limit of F, it is enough to show that its image under the Yoneda embedding $h^\bullet$ is the limit of $h^\bullet \circ F$ (Exercise 4.12). But since the Yoneda embedding preserves products and equalizers, it is enough to show that the analogue of (20) with $\mathcal{C}$ replaced by $\mathrm{Presh}(\mathcal{C})$ and F replaced by $h^\bullet \circ F$, has equalizer equal to the limit of $h^\bullet \circ F$. But this has been already observed in (17). $\square$

Remark 4.24. It is easy to check that if a category has binary products $X_1 \times X_2$, then it has all finite products $X_1 \times \cdots \times X_n$, except possibly for the empty product (i.e., the terminal object). Thus, by the above theorem, to check that a category is finitely complete, it is enough to check that it has equalizers, binary products, and a final object. A similar remark applies to check when a category is finitely cocomplete.

Exercise 4.25. Show that monomorphisms, epimorphisms, limits, colimits etc. all respect an equivalence of categories (with 'if and only if' statements).

5. Lecture 5 – Adjoint functors, tensor products

So far, we were ambiguous about whether *Ring* stood for the category of rings or that of commutative rings. We now disambiguate it, and reserve *Ring* for the category of not necessarily commutative rings (but associative, and with 1), and *CRing* for the category of commutative rings henceforth.

5.1. **Adjoint functors.** We continue with the convention that, unless otherwise stated, any category that we will encounter is locally small, though we will make an exception for presheaf categories of categories we work with.

Definition 5.1. (i) A functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ is said to be left adjoint to a functor $G : \mathcal{D} \rightsquigarrow \mathcal{C}$, or equivalently G is said to be right adjoint to F, if we have bijections

$$\text{adj} = \text{adj}_{X,Y} : \text{Mor}_{\mathcal{D}}(F(X), Y) \to \text{Mor}_{\mathcal{C}}(X, G(Y)) \tag{21}$$

for all $X \in \text{Ob}\,\mathcal{C}$ and $Y \in \text{Ob}\,\mathcal{D}$, that are natural (i.e., functorial) in X and Y, as captured by the following self-explanatory diagram for all $f : X' \to X$ in $\mathcal{C}$ and $g : Y \to Y'$ in $\mathcal{D}$:

$$\begin{array}{ccc} \text{Mor}_{\mathcal{D}}(F(X), Y) & \xrightarrow{\text{adj}_{X,Y}} & \text{Mor}_{\mathcal{C}}(X, G(Y)) \\ {\scriptstyle g\circ -\circ F(f)}\downarrow & & \downarrow{\scriptstyle G(g)\circ -\circ f} \\ \text{Mor}_{\mathcal{D}}(F(X'), Y') & \xrightarrow{\text{adj}_{X',Y'}} & \text{Mor}_{\mathcal{C}}(X', G(Y')) \end{array} . \tag{22}$$

In other words, if we have a natural isomorphism [6] of functors $\mathcal{C}^{op} \times \mathcal{D} \rightsquigarrow Set$,

$$adj : \text{Mor}_{\mathcal{D}}(F(-), -) \to \text{Mor}_{\mathcal{C}}(-, G(-)). \tag{23}$$

(ii) Such a natural isomorphism is called an adjunction between F and G.

Lemma 5.2. *(Uniqueness of adjoint functors) Any two functors $F, F' : \mathcal{C} \rightsquigarrow \mathcal{D}$ that are left adjoint to a given functor $G : \mathcal{D} \rightsquigarrow \mathcal{C}$ are naturally isomorphic to each other. A similar assertion holds, with 'left' replaced by 'right'.*

Proof. If $F, F' : \mathcal{C} \rightsquigarrow \mathcal{D}$ are both left adjoint to $G : \mathcal{D} \rightsquigarrow \mathcal{C}$, then we have identifications

$$\text{Mor}_{\mathcal{D}}(F(X), -) \cong \text{Mor}_{\mathcal{C}}(X, G(-)) \cong \text{Mor}_{\mathcal{D}}(F'(X), -)$$

for all $X \in \text{Ob}\,\mathcal{C}$, that are furthermore natural in X. By the Yoneda lemma (the $h_\bullet$ version), we have isomorphisms $F'(X) \to F(X)$ that are natural in $X \in \text{Ob}\,\mathcal{C}$, or in other words, a natural isomorphism $F \to F'$. This proves the assertion for left adjoint functors, and the assertion for right adjoint functors is proved similarly. □

[6] Remember that 'natural isomorphism' is a technical term defined precisely in Lecture 2.

Example 5.3. (i) Consider the functor $Forget : AbGrp \rightsquigarrow Set$, and the functor $FreeAb : Set \rightsquigarrow AbGrp$ that sends S to the free abelian group on S, denoted $FreeAb(S)$, which is a free abelian group containing S as a basis. [7] [8]

Since $S \subset FreeAb(S)$ is a basis, restriction to S gives, for each abelian group B, a bijection

$$\mathrm{Hom}_{AbGrp}(FreeAb(S), B) \to \mathrm{Hom}_{Set}(S, B) = \mathrm{Hom}_{Set}(S, Forget(B)).$$

It is easy to see that these bijections are natural in the sets S and the groups B, giving an adjunction realizing $S \rightsquigarrow FreeAb(S)$ as left adjoint to $Forget$.

(ii) Similarly, the functor $Forget : R\text{-}Mod \rightsquigarrow Set$ has a left adjoint that sends S to the free R-module on S, and in particular the functor $Forget : Vec_k \rightsquigarrow Set$ has a left adjoint that sends S to the free k-vector space on S.

(iii) The functor $Forget : Grp \rightsquigarrow Set$ has a left adjoint, sending a set S to the free group $Free(S)$ on S, reviewed in Subsection 5.3 below.

(iv) The functor $Forget : CRing \rightsquigarrow Set$ has a left adjoint, sending a set S to the polynomial ring $\mathbb{Z}[x_s \mid s \in S]$ (if you don't know what $\mathbb{Z}[x_s \mid s \in S]$ means, make the obvious guess; also, fill in the morphisms).

Indeed, any ring homomorphism $\mathbb{Z}[x_s \mid s \in S] \to R$ is completely determined by where the "free variables" x_s map to, so for any commutative ring R, we have a bijection:

$$\mathrm{Mor}_{CRing}(\mathbb{Z}[x_s \mid s \in S], R) \to \mathrm{Mor}_{Set}(S, R) = \mathrm{Mor}_{Set}(S, Forget(R)), \tag{24}$$

that sends φ to $(s \mapsto \varphi(x_s))$, whose inverse sends f to the unique $\varphi \in \mathrm{Mor}_{CRing}(\mathbb{Z}[x_s \mid s \in S], R)$ such that $\varphi(x_s) = f(s)$ for each $s \in S$. Show that (24) is indeed functorial in S and R, and hence gives the desired adjunction.

If we use R-algebras instead of $CRing$, something similar works, with $R[x_s \mid s \in S]$ instead of $\mathbb{Z}[x_s \mid s \in S]$.

The examples (i), (ii), (iii) and (iv) are examples of an informal principle called the 'free-forgetful adjunction': often a "forgetful functor" has a *left adjoint* provided by a "free construction". Similarly, e.g., we have a notion of "free monoids on S", a left adjoint to the forgetful functor from the category of monoids to Set (see Subsection 5.3 below).

(v) Consider $Forget : Top \rightsquigarrow Set$. Consider two functors $F, G : Set \rightsquigarrow Top$, where $F(S)$ is S with the discrete topology, and $G(S)$ is S with the indiscrete topology (as usual, fill in the morphisms). Then show that F is left adjoint to $Forget$, while G is right adjoint to $Forget$.

---

[7] More precisely, $FreeAb(S)$ is the group of formal sums $\sum_{s \in S} a_s s$, with $a_s \in \mathbb{Z}$ for each $s \in S$ and $a_s = 0$ for all but finitely many s, where the 's' of $a_s s$ is a formal symbol corresponding to $s \in S$. The inclusion $\iota_S : S \hookrightarrow FreeAb(S)$ that realizes S as a subset of $FreeAb(S)$ is defined by viewing each $s \in S$ as the element $\sum \delta_{s,s'} s' \in FreeAb(S)$, where $\delta_{s,s'}$ equals 1 if $s = s'$ and zero otherwise. Thus, $S \subset FreeAb(S)$ forms a basis.

[8] Recall, since this only defines the functor at the level of objects, you should mentally fill in how it 'should' be defined at the level of morphisms.

(vi) A left adjoint to the inclusion functor $AbGrp \rightsquigarrow Grp$ is the abelianization functor $Grp \rightsquigarrow AbGrp$ that sends G to its abelianization $G^{ab} := G/[G,G]$ (as usual, define the abelianization functor for morphisms; this will mostly not be repeated in future): if G is a group and H is an abelian group, then any homomorphism $G \to H$ factors as the composite of the abelianization map $G \to G^{ab}$ with a homomorphism $G^{ab} \to H$, and this induces a bijection

$$\mathrm{Mor}_{Grp}(G,H) \to \mathrm{Mor}_{AbGrp}(G^{ab},H).$$

(vii) Show that a left adjoint to the inclusion functor from the category of fields to that of integral domains, is given by the functor that takes any integral domain to its quotient field. See this by proving the following: any homomorphism from an integral domain R to a field factors uniquely through the obvious inclusion $R \hookrightarrow K$, where K is the quotient field of R.

(viii) A left adjoint to the inclusion functor from the category of Hausdorff topological spaces to Top, is given by the "Hausdorffization" functor, sending each topological space X to its "maximal Hausdorff quotient". The point is that any continuous map from X to a Hausdorff topological space factors uniquely through this quotient (show that this gives the desired adjunction).

A left adjoint to the inclusion functor from the category of compact Hausdorff spaces to that of all topological spaces, is given by the Stone-Čech compactification: any continuous map from X to a compact Hausdorff space factors uniquely through the Stone-Čech compactification (show that this gives the desired adjunction).

(ix) Let $Ring$ be the category of rings with 1, and Rng the category of rings without 1. The forgetful functor $Ring \rightsquigarrow Rng$ has a left adjoint given by "adjoining the identity": if $R \in \mathrm{Ob}\, Rng$, adjoining identity involves considering a ring whose underlying abelian group is $R^{\wedge} := \mathbb{Z} \oplus R$, where $1 \in \mathbb{Z} \subset R^{\wedge}$ acts as a multiplicative identity, and where the multiplication on $R \subset \mathbb{Z} \oplus R$ coincides with the multiplication coming from the Rng-structure on R. Again, the point is that any homomorphism in Rng from R to a ring R' with 1 factors uniquely through the inclusion $R \subset R^{\wedge}$ (show that this gives the desired adjunction).

(x) $\mathbb{Z}$ and $\mathbb{R}$ are directed sets, and hence can be viewed as categories as in Lecture 4: there is exactly one morphism $a \to b$ if $a \leqslant b$, and none otherwise. Then the inclusion functor $\mathbb{Z} \rightsquigarrow \mathbb{R}$ has:

- a left adjoint given has the ceiling function $x \mapsto \lceil x \rceil :=$ the smallest integer $\geqslant x$: indeed, for $a \in \mathbb{R}$ and $b \in \mathbb{Z}$, $a \leqslant b$ if and only if $\lceil a \rceil \leqslant b$; and
- a right adjoint given by the floor function $x \mapsto \lfloor x \rfloor :=$ the largest integer $\leqslant x$: indeed, for $a \in \mathbb{Z}$ and $b \in \mathbb{R}$, $a \leqslant b$ if and only if $a \leqslant \lfloor b \rfloor$.

(xi) We will see more examples, such as:

- Hom-tensor adjointness, $- \otimes_R N$ is left adjoint to $\mathrm{Hom}_R(N,-)$.
- Sending $V \in \mathrm{Ob}\, Vec_k$ to its tensor algebra $T(V)$, is a left adjoint to:

$$Forget : \text{Not necessarily commutative } k\text{-algebras} \rightsquigarrow Vec_k.$$

5.2. **Adjointness and commutativity with colimits/limits.** When we discuss tensor products, we will need to prove that tensor products commute with taking direct sums as well as with taking cokernels: $(M_1 \oplus M_2) \otimes_R N \cong M_1 \otimes_R N \oplus M_2 \otimes_R N$, and $\operatorname{coker}(M_1 \to M_2) \otimes_R N \cong \operatorname{coker}(M_1 \otimes_R N \to M_2 \otimes_R N)$. Both of these are special cases of the assertion that $- \otimes_R N$ commutes with colimits: note that a direct sum is a coproduct and hence a colimit, while a cokernel of f, being the coequalizer of f and the 0 map, is a colimit as well. This motivates the following definition:

Definition 5.4. A functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ is said to be continuous if it commutes with small limits, and cocontinuous if it commutes with small colimits.

That $- \otimes_R N$ commutes with small colimits, will be a consequence of its being left adjoint to some functor ($\operatorname{Hom}_R(N, -)$ in this case), so that the following proposition will apply:

Proposition 5.5. *Let $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ be a left adjoint to a functor $G : \mathcal{D} \rightsquigarrow \mathcal{C}$. Then F is cocontinuous, and G is continuous.*

Proof. We will prove the result for F; the result for G is analogous.

Suppose $(X, (\iota_j)_j)$ is a colimit of $H : J \rightsquigarrow \mathcal{C}$: this is the same as saying that for all $Y \in \operatorname{Ob} \mathcal{C}$, setting $X_j := H(j)$ for $j \in \operatorname{Ob} J$, we have a bijection:

$$\operatorname{Mor}_{\mathcal{C}}(X, Y) \overset{(-\circ\iota_j)_j}{\to} \Big\{ (f_j)_j \in \prod_{j \in \operatorname{Ob} J} \operatorname{Mor}(X_j, Y) \mid \forall\, h : i \to j \text{ in } J, \text{ we have } f_i = f_j \circ H(h) \Big\}.$$

Let adj be the adjunction between F and G, as in (23). Then for $Z \in \operatorname{Ob} \mathcal{D}$, we have bijections:

$$\begin{aligned} \operatorname{Mor}_{\mathcal{D}}(F(X), Z) &\overset{\text{adj}}{\to} \operatorname{Mor}_{\mathcal{C}}(X, G(Z)) \\ &\overset{(-\circ\iota_j)_j}{\to} \Big\{ (f_j)_j \in \prod_{j \in \operatorname{Ob} J} \operatorname{Mor}_{\mathcal{C}}(X_j, G(Z)) \mid \forall\, h : i \to j \text{ in } J, \text{ we have } f_i = f_j \circ H(h) \Big\} \\ &\overset{\text{adj}}{\to} \Big\{ (g_j)_j \in \prod_{j \in \operatorname{Ob} J} \operatorname{Mor}_{\mathcal{D}}(F(X_j), Z) \mid \forall\, h : i \to j \text{ in } J, \text{ we have } g_i = g_j \circ F(H(h)) \Big\}, \end{aligned}$$

where the last map is justified by the commutativity of the following diagram, for $h : i \to j$ in J:

$$\begin{array}{ccccc} f_j \in & \operatorname{Mor}_{\mathcal{C}}(X_j, G(Z)) & \xrightarrow{-\circ H(h)} & \operatorname{Mor}_{\mathcal{C}}(X_i, G(Z)) & \ni f_i \\ \downarrow & \uparrow{\scriptstyle \text{adj}} & & \uparrow{\scriptstyle \text{adj}} & \downarrow \\ g_j \in & \operatorname{Mor}_{\mathcal{D}}(F(X_j), Z) & \xrightarrow{-\circ F(H(h))} & \operatorname{Mor}_{\mathcal{C}}(F(X_i), Z) & \ni g_i \end{array},$$

which in turn follows from the naturality of adj (see (22), rotated 90 degrees anticlockwise).

To conclude from here that $(F(X), F(\iota_j)_j)$ is a colimit of $F \circ H : J \rightsquigarrow \mathcal{D}$, it is enough to show that the composite of the above chain of bijections is given by $(- \circ F(\iota_j))_j$. This is

equivalent to showing that the following diagram commutes for each $j \in \operatorname{Ob} J$:

$$\begin{array}{ccc} \operatorname{Mor}_{\mathcal{D}}(F(X), Z) & \xrightarrow{\text{adj}} & \operatorname{Mor}_{\mathcal{C}}(X, G(Z)) \\ {\scriptstyle -\circ F(\iota_j)}\downarrow & & \downarrow{\scriptstyle -\circ \iota_j} \\ \operatorname{Mor}_{\mathcal{D}}(F(X_j), Z) & \xrightarrow{\text{adj}} & \operatorname{Mor}_{\mathcal{C}}(X_j, G(Z)) \end{array},$$

which again follows from the naturality of adj (see (22)). □

We end this discussion with an exercise which might as well belong to an earlier subsection.

Exercise 5.6. This exercise introduces the units/counits of an adjunction, and explains how to describe the adjunction maps $\operatorname{adj}_{X,Y}$ in terms of these. Let F be left adjoint to G, and let an adjunction as in (23) between them be given, so we have the bijections $\operatorname{adj}_{X,Y}$ (see (21)). For all $X \in \operatorname{Ob}\mathcal{C}$ and $Y \in \operatorname{Ob}\mathcal{D}$, define

$$\eta_X : X \to G(F(X)) \qquad \text{and} \qquad \epsilon_Y : F(G(Y)) \to Y$$

as follows:

- $\eta_X = \operatorname{adj}_{X,F(X)}(\operatorname{id}_{F(X)}) \in \operatorname{Mor}_{\mathcal{C}}(X, G(F(X)))$ (see (21), apply it with $Y = F(X)$).
- $\epsilon_Y = \operatorname{adj}^{-1}_{G(Y),Y}(\operatorname{id}_{G(Y)}) \in \operatorname{Mor}_{\mathcal{D}}(F(G(Y)), Y)$ (see (21), apply it with $X = G(Y)$).

Show that $\eta = (\eta_X)_{X \in \operatorname{Ob}\mathcal{C}}$ is a natural transformation $id_{\mathcal{C}} \to G \circ F$, where $\operatorname{id}_{\mathcal{C}}$ is the identity functor $\mathcal{C} \rightsquigarrow \mathcal{C}$; one calls η the unit of the adjunction (23). Similarly $\epsilon = (\epsilon_Y)_{Y \in \operatorname{Ob}\mathcal{D}}$ is a natural transformation $F \circ G \to \operatorname{id}_{\mathcal{D}}$, called the counit of the adjunction.

Show that η and ϵ give ways to describe adj: namely, $\operatorname{adj}_{X,Y} : \operatorname{Mor}_{\mathcal{D}}(F(X), Y) \to \operatorname{Mor}_{\mathcal{D}}(X, G(Y))$ is given by $g \mapsto G(g) \circ \eta_X$, and the inverse bijection $\operatorname{adj}^{-1}_{X,Y} : \operatorname{Mor}_{\mathcal{D}}(X, G(Y)) \to \operatorname{Mor}_{\mathcal{D}}(F(X), Y)$ is given by $f \mapsto \epsilon_Y \circ F(f)$.

Example 5.7. The forgetful functors $AbGrp \rightsquigarrow Set$ and $R\text{-}Mod \rightsquigarrow Set$ do not respect coproducts: they take a direct sum to a set-theoretic product rather than to a disjoint union. One can show that the forgetful functors $Grp \rightsquigarrow Set$ and $CRing \rightsquigarrow Set$ do not respect coproducts either (the former can be shown using material later in this chapter). Therefore, these functors do not have a right adjoint, unlike the forgetful functor $Top \rightsquigarrow Set$ (Example 5.3(v)). Recall, though, that these functors all had a left adjoint (see Example 5.3).

Similarly, the inclusion functor $AbGrp \rightsquigarrow Grp$ does not respect coproducts: the material later in this lecture will show that given two abelian groups A and B, their coproduct in the category Grp (which is called a free product of A and B and denoted $A * B$) is typically nonabelian and hence different from their coproduct $A \oplus B$ in $AbGrp$. For instance, this is the case when $A = B = \mathbb{Z}$. Thus, this functor does not have a right adjoint either, though we saw that it has a left adjoint given by abelianization.

5.3. **Free groups.** Let us now discuss the free group on a set S, denoted $Free(S)$. Some of this subsection will be informal (e.g., the definition of the equivalence relation $\sim$ in Construction 5.10 below), and at least two of the proofs will be skipped (Proposition 5.11 and Lemma 5.15). If you want a more detailed treatment, please look up a reference, such as the book of Dummit and Foote, or Section 2.2.2 of "Geometric Group Theory: An Introduction" by Clara Löh, about which I learnt from Radhika Gupta. A different treatment can be seen in Serge Lang's "Algebra", where he defines $Free(S)$ to be a group that satisfies the universal property we expect from it (see Lemma 5.13 below), first showing the existence of such a group using an argument due to Tits, and then giving an approach (or two) showing that this group can be described as in more conventional treatments (as followed in the discussion below).

Definition 5.8. Let S be a set. The free monoid on S is the set $FreeMon(S)$ of all sequences of elements of S, made into a monoid via concatenation: a typical sequence $(s_1, \ldots, s_n) \in FreeMon(S)$ is called a word and written $s_1 s_2 \ldots s_n$, and the product of $s_1 s_2 \ldots s_n$ and $t_1 t_2 \ldots t_m$ is written $s_1 s_2 \ldots s_n t_1 t_2 \ldots t_m$. This includes the empty sequence, which is written 1 and is the identity of $FreeMon(S)$.

There is an obvious inclusion $S \hookrightarrow FreeMon(S)$, sending $s \in S$ to the singleton sequence with just s in it.

Exercise 5.9. (i) Show that restriction with respect to the obvious inclusion $S \subset FreeMon(S)$ induces, for any monoid M, a bijection

$$\mathrm{Mor}_{Monoid}(FreeMon(S), M) \to \mathrm{Mor}_{Set}(S, M). \tag{25}$$

(ii) Use (i) to extend $S \mapsto FreeMon(S)$ to a functor $FreeMon : Set \rightsquigarrow Monoid$, and show that the maps of (25), as the set S and the monoid M vary, realize $FreeMon$ as a left adjoint to the forgetful functor $Forget : Monoid \rightsquigarrow Set$.

Construction 5.10. Let S be a set. Introduce a symbol s^{-1} for all $s \in S$, and let $T = S \sqcup S^{-1}$, where $S^{-1} = \{s^{-1} \mid s \in S\}$. Given $w_1, w_2 \in FreeMon(T)$, declare $w_1 \sim w_2$ if they can be obtained from each other by a finite sequence of insertions or deletions of words of the form xx^{-1} or $x^{-1}x$, where $x \in S \subset T$. This is clearly an equivalence relation, so we may form $FreeMon(T)/\sim$.

The following proposition is not difficult at all, but I don't want to write out the details:

Proposition 5.11. *In the setting of Construction 5.10, the multiplication on $FreeMon(T)$ descends to a well-defined multiplication on $FreeMon(T)/\sim$, making it into a group.*

Proof. Omitted. Please look it up somewhere, say Section 2.2.2 of Clara Löh's book. □

Definition 5.12. Let S be a set. Then we define $Free(S)$ to be the group whose underlying set is $FreeMon(T)/\sim$ as in Construction 5.10, with multiplication induced from that in $FreeMon(T)$ (as justified by Proposition 5.11). Let $\iota_S : S \to Free(S)$ be the obvious map.

Recall that a group G is generated by $S' \subset G$ if the smallest subgroup of G containing S' equals G. Equivalently, if every element of G can be written as a product $s_1^{a_1} \dots s_n^{a_n}$, where $s_i \in S'$ and $a_i \in \mathbb{Z}$. Clearly, the image of S in $Free(S)$ generates $Free(S)$.

To construct $S \rightsquigarrow Free(S)$ as a functor, i.e., to define it at the level of morphisms, we will use the following lemma.

Lemma 5.13. *Restriction to S (i.e., $- \circ \iota_S$) defines, given any group G, a bijection*

$$\mathrm{Mor}_{Grp}(Free(S), G) \to \mathrm{Mor}_{Set}(S, G) = \mathrm{Mor}_{Set}(S, Forget(G)). \tag{26}$$

Proof. The injectivity follows from the fact that $\iota_S(S)$ generates $Free(S)$. For surjectivity, given $\varphi \in \mathrm{Mor}_{Set}(S, Forget(G))$, note that the universal property of $FreeMon(T)$ gives us a unique monoid homomorphism $FreeMon(T) \to G$ that, for each $s \in S$, sends s to $\varphi(s)$ and s^{-1} to $\varphi(s)^{-1}$. It is clear that this homomorphism respects the equivalence relation $\sim$, and hence descends to a monoid homomorphism $FreeMon(T)/\sim \to G$, which is the same as a group homomorphism $Free(S) \to G$, whose composite with ι_S is $\varphi : S \to G$. □

Corollary 5.14. *The map $\iota_S : S \to Free(S)$ is injective.*

Proof. Given $s_1, s_2 \in S$ with $s_1 \neq s_2$, there exists $\varphi \in \mathrm{Mor}_{Set}(S, \mathbb{Z})$ with $\varphi(s_1) \neq \varphi(s_2)$, and hence, by Lemma 5.13, a homomorphism $\psi : Free(S) \to \mathbb{Z}$ with $\psi \circ \iota_S(s_1) \neq \psi \circ \iota_S(s_2)$, forcing $\iota_S(s_1) \neq \iota_S(s_2)$. □

Lemma 5.13 lets us define $S \rightsquigarrow Free(S)$ as a functor: given $f : S \to S'$, define $Free(f) : Free(S) \to Free(S')$ by taking $G = Free(S')$ in (26), and letting $Free(f)$ be the inverse image under that bijection of

$$\iota_{S'} \circ f : S \to Free(S') = G.$$

It is easy to check that $Free(f_2 \circ f_1) = Free(f_2) \circ Free(f_1)$ and that $Free(\mathrm{id}_S) = \mathrm{id}_{Free(S)}$.

We will probably not use the lemma below, but it tells us what elements in $Free(S)$ look like, justifying the use of the word 'free'.

Lemma 5.15. *Every equivalence class in $Free(S) = FreeMon(T)/\sim$ contains a unique reduced word, by which we mean a word that does not contain any subsequence of the form xx^{-1} or $x^{-1}x$, with $x \in S$. Thus, every element of $Free(S)$ can be uniquely written in the form*

$$s_1^{a_1} \dots s_n^{a_n},$$

where $s_i \in S$ and $a_i \in \mathbb{Z}\backslash\{0\}$ for each i, and $s_i \neq s_{i+1}$ for all $1 \leqslant i \leqslant n-1$ (note that n is allowed to be 0, in which case the above element is the identity element of $Free(S)$).

Proof. The proof of the first assertion is omitted, please look up some source such as Dummitt and Foote. The second assertion is a restatement of the first. □

5.4. **Presentation of groups by generators and relations.** A presentation of a group G, or a description of G using generators and relations, is a triple (S, R, h) consisting of a set S, a subset $R \subset Free(S)$, and a surjective homomorphism $h : Free(S) \to G$ such that $\ker(h) \subset Free(S)$ is the normal subgroup of $Free(S)$ generated by R, i.e., the smallest normal subgroup of $Free(S)$ that contains R.

Note that given such a presentation, $h(S) \subset G$ generates G, since $h : Free(S) \to G$ is surjective, and S generates $Free(S)$. One might often suppress h from the notation and say that G has a presentation $\langle S|R\rangle$ or that $G \cong \langle S|R\rangle$.

Lemma 5.16. *Any group G has a presentation.*

Proof. Let G be any group and let $S \subset G$ be a set of generators. Consider $Free(S)$; one has inclusion morphisms $\iota_S : S \hookrightarrow Free(S)$ and $f : S \hookrightarrow G$, but this should not cause confusion (between, e.g., $s_1^3 s_2$ viewed as an element of $Free(S)$, and $s_1^3 s_2$ viewed as an element of G; the latter may be trivial but the former is never so). Corresponding to the latter map, $f : S \hookrightarrow G$, (26) gives us a homomorphism $h : Free(S) \to G$ sending $\iota_S(s)$ to $f(s)$. Since S was chosen as a set of generators for G, it follows that $h : Free(S) \to G$ is surjective, i.e., every group is a quotient of a free group. One can then take R to be any set of generators for $\ker(h)$, for instance $\ker(h)$ itself. □

Example 5.17. It is easy to see that for any group G, $G \cong \langle G|xyz^{-1} \mid z = xy \in G\rangle$: if $G_0 \subset Free(G)$ is the normal subgroup generated by the xyz^{-1} with $z = xy \in G$, then restriction via $G \hookrightarrow Free(G)$ gives bijections:

$$\mathrm{Mor}_{Grp}(Free(G)/G_0, H) \to \{f \in \mathrm{Mor}_{Set}(G, H) \mid f(x)f(y) = f(z)\,\forall\, z = xy \in G\} = \mathrm{Mor}_{Grp}(G, H).$$

Example 5.18.

(i) For $G = \mathbb{Z}/n\mathbb{Z}$, one can take $S = \{a\}$ to be a singleton, and $R = \{a^n\}$, so $G \cong \langle a|a^n\rangle = \langle a, a^2|a^n, a^{2n}\rangle$.

(ii) A dihedral group D_n of order $2n$ has a presentation

$$D_n = \langle r, f \mid r^n, f^2, (rf)^2\rangle,$$

which we may also write as $\langle r, f \mid r^n = f^2 = (rf)^2 = 1\rangle$ or as $\langle r, f \mid r^n = f^2 = 1, frf^{-1} = r^{-1}\rangle$ (note that there are multiple irredundant ways to present any group).

Here, the use of the letters 'r' and 'f' is motivated by the usual realization of D_n as a group of symmetries of the Euclidean space $\mathbb{R}^2$, where 'the Euclidean space $\mathbb{R}^2$' means 'the vector space $\mathbb{R}^2$ together with its standard inner product': r stands for an anticlockwise rotation of $2\pi/n$ about the origin, and f stands for a flip, or a reflection, across the line through the origin that makes an angle of π/n with the x-axis. r and f in this picture could be more general (exercise: try to describe the other possible r and f in the picture).

(iii) The symmetric group S_n on n letters can be shown to have a presentation

$$\langle s_1, \ldots, s_{n-1}|s_i^2 = 1\,\forall\, 1 \leqslant i \leqslant n-1, s_i s_j = s_j s_i\,\forall\, j \neq i\pm 1, s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}\,\forall 1 \leqslant i \leqslant n-2\rangle.$$

Let us sketch how to show this. Writing W_n for the group with the above presentation, first, it is easy to see that sending s_i to the transposition $(i, i+1)$ gives

a homomorphism $W_n \to S_n$: this is because it is easy to compute that the above relations are satisfied when each s_i replaced by $(i, i+1)$. This homomorphism $W_n \to S_n$ is surjective, since the $(i, i+1)$ generate S_n. Therefore, it is enough to show that $\#W_n = n!$.

We will do this by induction. The analogous map $W_{n-1} \to S_{n-1}$ is an isomorphism by induction, so $\#W_{n-1} = (n-1)!$. Therefore, it is enough to show that $\#(W_n/W_{n-1}) = n$, which follows if we show that any right coset of W_{n-1} in W_n contains a representative from $\{1, s_{n-1}, s_{n-2}s_{n-1}, \ldots, s_1 \ldots s_{n-1}\}$ (we assume this inductively, with n replaced by $n-1$). For this, given $w \in W_n \backslash W_{n-1}$ expressed as a product of the s_i, one first manipulates this product using the relations so as to have exactly one copy of s_{n-1}. This implies that $W_n = W_{n-1} \sqcup W_{n-1}s_{n-1}W_{n-1}$. Therefore:

$$\begin{aligned} W_n &= W_{n-1} \sqcup W_{n-1}s_{n-1}W_{n-1} \\ &= W_{n-1} \sqcup (\{1\} \sqcup \{s_j \ldots s_{n-2} \mid 1 \leqslant j \leqslant n-2\})\, W_{n-2}s_{n-1}W_{n-1} \\ &= W_{n-1} \sqcup s_{n-1}W_{n-1} \sqcup \{s_j \ldots s_{n-2} \mid 1 \leqslant j \leqslant n-2\}s_{n-1}W_{n-1}, \end{aligned}$$

where in the second step we used the induction hypothesis to get coset representatives for W_{n-1}/W_{n-2}, and in the third step we used the fact that W_{n-2} can be commuted past s_{n-1}. This gives the desired set of coset representatives.

(iv) $PSL_2(\mathbb{Z})$ has a presentation $\langle a, b \mid a^2, b^3 \rangle$, where a maps to $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and b to $\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$. This is not obvious; you can look up if you are interested but don't like a challenge at this point.

(v) The free abelian group on S has a presentation $\langle S | \{aba^{-1}b^{-1} \mid a, b \in S\} \rangle$ (exercise).

5.5. **Colimits in the category of groups.**

- If $G_i = \langle S_i | R_i \rangle$ for all $i \in I$, let

$$G := \langle S := \bigsqcup_{i \in I} S_i \mid R := \bigsqcup_{i \in I} R_i \rangle.$$

Then it is immediate that for each i, we get a map $\iota_i : G_i \to G$ (sending the image of s_i in G_i to the image of s_i in G), such that the images of the ι_i generate G. The map

$$\mathrm{Hom}(G, H) \overset{(-\circ\iota_i)_{i\in I}}{\to} \prod_i \mathrm{Hom}(G_i, H)$$

is bijective: the injectivity follows since every element of S (or rather, its image in G) belongs to the image of some G_i, while for surjectivity, use that given $\varphi_i \in \mathrm{Hom}(G_i, H)$ for each i, sending $s_i \in S_i \subset S$ to $\varphi_i(s_i)$ for each i satisfies each relation in R and hence gives a well-defined homomorphism $\varphi : G \to H$ with $\varphi|_{G_i} = \varphi_i$ for each i. Therefore, $(G, (\iota_i)_i)$ is a coproduct of the G_i. Note that each ι_i is injective, so we may think of the G_i as subgroups of G, that together generate G. The binary

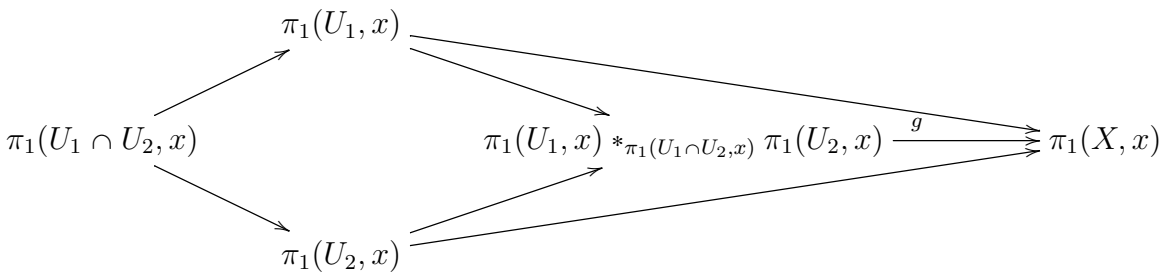
coproduct $G_1 \sqcup G_2$ is also denoted $G_1 * G_2$, and called the free product of G_1 and G_2.

- More generally, let $F : J \rightsquigarrow Grp$ be a functor, with J a small category. Write $G_j = F(j)$, and form the coproduct $(G := \bigsqcup_{j \in \mathrm{Ob}\, J} G_j, \iota_j : G_j \to G)$. Each ι_j is a monomorphism, and can hence be viewed as an inclusion, so we regard the G_j as subgroups of G. Then a colimit of F is given as the quotient of G by the normal subgroup generated by
$$\{g_j(F(f)(g_i))^{-1} \mid f : i \to j \text{ in } J, g_j \in G_j, g_i \in G_i\}.$$
- A special case is when we have $\{G_i \mid i \in I\}$ indexed by a set I, and a group A mapping to each G_i. This realizes an obvious functor $F : J \rightsquigarrow G$, where $\mathrm{Ob}\, J = I \sqcup \{*\}$, whose colimit G is called the amalgamated product of the G_i. Explicitly, check that this colimit G is the quotient of $\bigsqcup_i G_i$ by the normal subgroup generated by
$$\{\varphi_i(a)\varphi_j(a)^{-1} \mid i, j \in I, a \in A\}.$$
Note that when $I = \{1, 2\}$ has two elements, the above is nothing but a pushout of $A \to G_1$ and $A \to G_2$: we will denote this by $G_1 *_A G_2$.

Example 5.19. The presentation of $PSL_2(\mathbb{Z})$ given in Example 5.18(iv) implies that $PSL_2(\mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z}) * (\mathbb{Z}/3\mathbb{Z})$. Note that, while $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are finite, their free product is the infinite "large looking" group $PSL_2(\mathbb{Z})$.

In topology, you will learn the Seifert-Van Kampen theorem:

Theorem 5.20 (Seifert-Van Kampen theorem)**.** *Suppose a topological space X is the union of two path connected subspaces U_1 and U_2, such that $U_1 \cap U_2$ is nonempty and path connected. Let $x \in U_1 \cap U_2$, so applying the functor π_1 to the inclusions $(U_1 \cap U_2, x) \hookrightarrow (U_i, x) \to (X, x)$ of pointed topological spaces give a commutative diagram:*



(see the explanation below the theorem). Then the map $g : \pi_1(U_1, x) *_{\pi_1(U_1 \cap U_2, x)} \pi_1(U_2, x) \to \pi_1(X, x)$ in the above diagram is an isomorphism, realizing $\pi_1(X, x)$ as the amalgamated product of $\pi_1(U_1, x)$ and $\pi_1(U_2, x)$ over $\pi_1(U_1 \cap U_2, x)$.

To explain the diagram in the above theorem, we used that the composite map $\pi_1(U_1 \cap U_2, x) \hookrightarrow \pi_1(U_i, x) \to \pi_1(X, x)$ is independent of $i \in \{1, 2\}$, being obtained by applying π_1 to the inclusion $(U_1 \cap U_2, x) \hookrightarrow (X, x)$; this is why, by the universal property of pushouts, we get the map $g : \pi_1(U_1, x) *_{\pi_1(U_1 \cap U_2, x)} \pi_2(U_2, x) \to \pi_1(X, x)$.

Thus, since $\pi_1(S^1, \{x\}) \cong \mathbb{Z}$, we get that the fundamental group of the "figure 8" space, a union of two circles intersecting at exactly one point, is $\mathbb{Z} * \mathbb{Z}$, which is just a free group on two generators.

6. Lecture 6 – tensor products over commutative rings

Correction: In Lecture 5, I defined a continuous functor as one that preserves all limits. This is wrong: a continuous functor is one that preserves all *small* limits. Similarly, a cocontinuous functor is one that preserves all small colimits.

6.1. **The definition of tensor products.**

Notation 6.1. Throughout today's lecture, R will denote a commutative ring. When $\mathcal{C} = R\text{-}Mod$, $\mathrm{Mor}_{\mathcal{C}}$ will be denoted by Hom_R.

Definition 6.2. (i) For modules $M_1, \ldots, M_r$ and L over a commutative ring R, write

$$\mathrm{Mult}_R(M_1, \ldots, M_r; L) = \{f : M_1 \times \cdots \times M_r \to L \mid f \text{ is } R\text{-multilinear}\},$$

where we recall that 'R-multilinear' means 'R-linear in each variable'. $\mathrm{Mult}_R(M_1, \ldots, M_r; L)$ has an obvious structure of an R-module, using the R-module structure on L.

(ii) $\mathrm{Mult}(M_1, \ldots, M_r; -) = \mathrm{Mult}_R(M_1, \ldots, M_r; -)$ [9] can be thought of either as a functor $R\text{-}Mod \rightsquigarrow R\text{-}Mod$, or as a functor $R\text{-}Mod \rightsquigarrow Set$. When $r = 2$, we write Bil_R for Mult_R.

(iii) A tensor product of $M_1, \ldots, M_r$ over R can be defined in one of the following two equivalent ways:

- It is an R-module M, together with an R-multilinear map $u : M_1 \times \cdots \times M_r \to M$, such that for each R-module L, the map

$$\mathrm{Hom}_R(M, L) \xrightarrow{-\circ u} \mathrm{Mult}_R(M_1, \ldots, M_r; L)$$

is a bijection (and hence an isomorphism of R-modules). In other words, $- \circ u$ is a natural isomorphism of functors $\mathrm{Hom}_R(M, \cdot) \to \mathrm{Mult}_R(M_1, \ldots, M_r; -)$.

- It is an R-module $M \in \mathrm{Ob}\, R\text{-}Mod$ that corepresents the functor $Mult_R(M_1, \ldots, M_r; -)$, together with a natural isomorphism $h_M \to \mathrm{Mult}_R(M_1, \ldots, M_r; -)$ of functors $R\text{-}Mod \rightsquigarrow Set$.

As you should be able to immediately see by now, a tensor product (M, u) of $M_1, \ldots, M_r$ is *uniquely* unique if it exists, so I am not even listing that as an exercise.

Exercise 6.3. (i) Convince yourself of the equivalence between the two ways of defining a tensor product of $M_1, \ldots, M_r$ (hint: see Remark 3.15 from Lecture 3: $u \in \mathrm{Mult}_R(M_1, \ldots, M_r; M)$ is what was informally referred to as the 'universal object').

(ii) (Easy) If (M, u) is a tensor product of $M_1, \ldots, M_r$, show that $u(M_1 \times \cdots \times M_r)$ spans M.

Notation 6.4. Instead of denoting a tensor product of $M_1, \ldots, M_r$ by (M, u), we will typically write $M_1 \otimes_R \cdots \otimes_R M_r$ in place of M, or even $M_1 \otimes \cdots \otimes M_r$ or $\bigotimes_{i=1}^{r} M_i$ when R is understood from the context, and write $m_1 \otimes \cdots \otimes m_r$ for $u(m_1, \ldots, m_r) \in M$.

[9] One uses the former when R is understood.

Note that this notation does encode information about u, since u is then given by $(m_1, \dots, m_r) \mapsto m_1 \otimes \cdots \otimes m_r$, and under it the defining property of the tensor product can be rephrased as follows: given a multilinear map $h : M_1 \times \cdots \times M_r \to L$, there exists a unique linear map $f : M_1 \otimes_R \cdots \otimes_R M_r \to L$ with the property that $f(m_1 \otimes \cdots \otimes m_r) = h(m_1, \dots, m_r)$ whenever $m_i \in M_i$ for all $1 \leqslant i \leqslant r$. This notation is well-defined, since the tensor product is uniquely unique. This is how we will mostly write the tensor product from now on.

Remark 6.5. Assume that $M_1 \otimes_R \cdots \otimes_R M_r$ exists (which we will prove soon).

(i) By Exercise 6.3(ii), while not every element of $M_1 \otimes_R \cdots \otimes_R M_r$ is of the form the $m_1 \otimes \cdots \otimes m_r$, every element of $M_1 \otimes_R \cdots \otimes_R M_r$ can be written as a finite linear combination (or equivalently, as a finite sum) of terms of the form $m_1 \otimes \cdots \otimes m_r$, which may be referred to as 'pure tensors'.

(ii) This has the consequence that every R-module homomorphism $M_1 \otimes_R \cdots \otimes_R M_r \to L$, for any R-module L, is uniquely determined by where it sends the $m_1 \otimes \cdots \otimes m_r$: this fact will be implicitly used in what follows, to automatically consider various uniqueness assertions as having been proved.

(iii) The multilinearity of the above map $M_1 \times \cdots \times M_r \to M_1 \otimes_R \cdots \otimes_R M_r$ means that the expression $m_1 \otimes \cdots \otimes m_r$ is multilinear in $m_1, \dots, m_r$, and that it vanishes when some m_i equals 0.

Here are some examples of tensor products we can already calculate.

Proposition 6.6. *(i) For any R-module N, a tensor product of R and N is given by the (obviously bilinear) multiplication map $R \times N \to N$, i.e., $(a, n) \mapsto an$, and a tensor product of N and R by the map $N \times R \to R$, $(n, a) \mapsto an$. In other words, using Notation 6.4, we have isomorphisms $R \otimes_R N \cong N \cong N \otimes_R R$, under which, for all $a \in R$ and $n \in N$, $a \otimes n \mapsto an \leftarrow\!\shortmid n \otimes a$.*

(ii) For any R-module N and ideal $I \subset N$, we have unique isomorphisms $R/I \otimes_R N \cong N/IN \cong N \otimes_R R/I$, such that for all $a \in R$ and $n \in N$, $\bar{a} \otimes n \mapsto \overline{an} \leftarrow\!\shortmid n \otimes \bar{a}$, where $\bar{a}$ and $\overline{an}$ are the images of a and an in R/I and N/IN (Exercise: translate this into the "(M, u)" notation).

Remark 6.7. Recall that IN is not $\{an \mid a \in I, n \in N\}$, but $\mathrm{Span}_R(\{an \mid a \in I, n \in N\})$.

Proof of Proposition 6.6. We have an R-module homomorphism $\mathrm{Bil}_R(R, N; L) \to \mathrm{Hom}_R(N, L)$, sending B to $n \mapsto B(1, n)$. This map is an isomorphism, since a two-sided inverse is readily checked to be the map sending $\varphi \in \mathrm{Hom}_R(N, L)$ to $(a, n) \mapsto a\varphi(n) = \varphi(an)$. This isomorphism being clearly functorial in L, the multiplication map $R \times N \to N$ is a tensor product for R and N. A similar argument applies to bilinear maps on $N \times R$, and (i) follows.

Now we come to (ii). Note that $(\bar{a}, n) \mapsto \overline{an}$ is a well-defined map $R/I \times N \to N/IN$, which is bilinear. Therefore, for any R-module L, $- \circ ((\bar{a}, n) \mapsto \overline{an})$ is an R-module map $\mathrm{Hom}_R(N/IN, L) \to \mathrm{Bil}_R(R/I, N; L)$, and it is enough to prove that this map is an

isomorphism. It is the left vertical arrow of the following diagram:

$$\begin{array}{ccc} \mathrm{Hom}_R(N/IN, L) & \xhookrightarrow{-\circ(N\to N/IN)} & \mathrm{Hom}_R(N, L) \\ \Big\downarrow{\scriptstyle -\circ((\bar{a},n)\mapsto\overline{an})} & & \Big\downarrow{\scriptstyle -\circ((a,n)\mapsto an)} \\ \mathrm{Bil}_R(R/I, N; L) & \xhookrightarrow{-\circ((a,n)\mapsto(\bar{a},n))} & \mathrm{Bil}_R(R, N; L) \end{array}.$$

It is clear that this diagram is commutative. The top rows are clearly injections, and hence will be viewed as inclusions. By (i), the right vertical arrow in the diagram is an isomorphism, so it is enough to show that it maps the submodule $\mathrm{Hom}_R(N/IN, L) \subset \mathrm{Hom}_R(N, L)$ onto the submodule $\mathrm{Bil}_R(R/I, N; L) \subset \mathrm{Bil}_R(R, N; L)$; or equivalently, that its inverse maps $\mathrm{Bil}_R(R/I, N; L)$ to $\mathrm{Hom}_R(N/IN, L)$.

$\mathrm{Bil}_R(R/I, N; L) \subset \mathrm{Bil}_R(R, N; L)$ is precisely the submodule consisting of the bilinear maps that vanish on $I \times N$. The inverse image of this submodule under the right vertical arrow is the submodule of $\mathrm{Hom}_R(N, L)$ that vanishes on $\{an \mid a \in I, n \in N\}$, or equivalently, on its span. But this span is precisely IN. In other words, the isomorphism $\mathrm{Bil}_R(R, N; L) \to \mathrm{Hom}_R(N, L)$ sends $\mathrm{Bil}_R(R/I, N; L)$ to the submodule of $\mathrm{Hom}_R(N, L)$ that vanishes on $IN \subset N$, which identifies with $\mathrm{Hom}_R(N/IN, L) \subset \mathrm{Hom}_R(N, L)$, via the top horizontal arrow, which was what we wanted to prove. □

Example 6.8. In particular:

- $R/I \otimes_R R/J \cong (R/J)/I(R/J) \cong R/(I + J)$.
- As an even more special case, it follows that $(\mathbb{Z}/m\mathbb{Z}) \otimes_{\mathbb{Z}} (\mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}/(m, n)\mathbb{Z}$, which is 0 if $(m, n) = 1$. Thus, for instance, $(\mathbb{Z}/2\mathbb{Z}) \otimes_{\mathbb{Z}} (\mathbb{Z}/3\mathbb{Z}) = 0$, showing that, unlike what we will see in Example 6.12 below, there can be a lot of 'collapsing' when one takes tensor products of non-free modules.

Lemma 6.9. *If $f_i : M_i \to N_i$ is an R-module homomorphism for $1 \leqslant i \leqslant r$, and both $\bigotimes_{i=1}^r M_i$ and $\bigotimes_{i=1}^r N_i$ exist, then there exists a unique R-module homomorphism*

$$\bigotimes_{i=1}^{r} f_i : \bigotimes_{i=1}^{r} M_i \to \bigotimes_{i=1}^{r} N_i$$

with the property that whenever $m_i \in M_i$ for all $1 \leqslant i \leqslant r$, we have

$$\Big(\bigotimes_{i=1}^{r} f_i\Big)(m_1 \otimes \cdots \otimes m_r) = f_1(m_1) \otimes \cdots \otimes f_r(m_r).$$

Moreover, this construction is functorial: stated informally, if $M_i \xrightarrow{f_i} N_i \xrightarrow{g_i} L_i$, and $\bigotimes_{i=1}^r ?_i$ exists for $? \in \{M, N, L\}$, then

$$\bigotimes_{i=1}^{r}(g_i \circ f_i) = \Big(\bigotimes_{i=1}^{r} g_i\Big) \circ \Big(\bigotimes_{i=1}^{r} f_i\Big), \qquad \bigotimes_{i=1}^{r} \mathrm{id}_{M_i} = \mathrm{id}_{\bigotimes_{i=1}^r M_i}.$$

Proof. The first assertion follows from the fact that $(m_1, \dots, m_r) \mapsto f_1(m_1) \otimes \cdots \otimes f_r(m_r)$ is R-multilinear, together with the defining property of the tensor product $\bigotimes_{i=1}^r M_i$. One can also obtain this by applying the Yoneda lemma to the natural transformation

$$\mathrm{Mult}_R(N_1, \dots, N_r, -) \to \mathrm{Mult}_R(M_1, \dots, M_r, -)$$

obtained by pulling back with respect to

$$(f_1, \dots, f_r) : M_1 \times \cdots \times M_r \to N_1 \times \cdots \times N_r.$$

The functoriality is immediate from the uniqueness (which should be understood to follow as explained in Remark 6.5(ii)). □

Proposition 6.10. *(i) Suppose $M \otimes_R N_i$ and $M_i \otimes_R N$ exist for each $i \in I$. Then $M \otimes_R (\bigoplus_i N_i)$ and $(\bigoplus_i M_i) \otimes_R N$ exist, and there exists a unique isomorphism*

$$M \otimes_R \Big(\bigoplus_{i\in I} N_i\Big) \to \bigoplus_{i\in I} M \otimes_R N_i, \qquad \Big(\textit{resp.,}\ \Big(\bigoplus_{i\in I} M_i\Big) \otimes_R N \to \bigoplus_{i\in I} M_i \otimes_R N\Big),$$

defined by the requirement that it send each $m \otimes (n_i)_i$ to $(m \otimes n_i)_i$, (resp., each $(m_i)_i \otimes n$ to $(m_i \otimes n)_i$).

(ii) Suppose a sequence of homomorphisms

$$M_1 \xrightarrow{f} M_2 \xrightarrow{g} M \to 0$$

of R-modules is exact, and that $M_1 \otimes_R N$ and $M_2 \otimes_R N$ exists. Then $M \otimes_R N$ exists, and the sequence

$$M_1 \otimes_R N \xrightarrow{f\otimes \mathrm{id}_N} M_2 \otimes_R N \xrightarrow{g\otimes \mathrm{id}_N} M \otimes_R N \to 0 \tag{27}$$

(where $f \otimes \mathrm{id}_N$ and $g \otimes \mathrm{id}_N$ are as in Lemma 6.9) is exact as well. A similar assertion applies to tensoring with N on the left.

Remark 6.11. (i) Please don't read Proposition 6.10(i) as just saying that $(\bigoplus_i M_i) \otimes_R N$ is isomorphic to $\bigoplus_i M_i \otimes_R N$: it is making the stronger statement that there is a very particular isomorphism $\bigoplus_i M_i \otimes_R N \to (\bigoplus_i M_i) \otimes_R N$, sending each $(m_i \otimes n)_i$ to $(m_i)_i \otimes n$.

(ii) Here is another way to express the condition that $\bigoplus_i M_i \otimes_R N \to (\bigoplus_i M_i) \otimes_R N$ takes $(m_i \otimes n)_i$ to $(m_i)_i \otimes n$. Namely, applying the functoriality of $- \otimes_R N$ to the inclusion $\iota_j : M_j \hookrightarrow \bigoplus_i M_i$ gives us a homomorphism $\iota_j \otimes \mathrm{id}_N : M_j \otimes_R N \to (\bigoplus_i M_i) \otimes_R N$, and the condition $(m_i \otimes n)_i \mapsto (m_i)_i \otimes n$ above is equivalent to saying that the $(\iota_i \otimes \mathrm{id}_N)_i$ together sum up to an isomorphism $\bigoplus_i M_i \otimes_R N \to (\bigoplus_i M_i) \otimes_R N$. Thus, Proposition 6.10(i) is essentially saying "$- \otimes_R N$ respects coproducts".

(iii) If it is not already obvious to you, please make sure you understand that (ii) of the proposition can be rephrased as follows. Suppose $f : M_1 \to M_2$ is a homomorphism of R-modules, and let $M := \mathrm{coker}(f)$. If $M_1 \otimes_R N$ and $M_2 \otimes_R N$ exist, then $M \otimes_R N$ exists as well, and can be described as follows: if $g : M_2 \to M$ is the obvious surjection, then (ii) of the proposition is telling us that $g \otimes \mathrm{id}_N : M_2 \otimes_R N \to M \otimes_R N$

has kernel equal to the image of $f \otimes \mathrm{id} : M_1 \otimes_R N \to M_2 \otimes_R N$, and that the resulting map

$$\mathrm{coker}(f \otimes \mathrm{id}_N : M_1 \otimes_R N \to M_2 \otimes_R N) \to \mathrm{coker}(f : M_1 \to M_2) \otimes_R N = M \otimes_R N$$

is an isomorphism. In other words, "tensoring respects cokernels".

(iv) The property in (ii) of the proposition will be referred to as saying that the "operations" [10] $- \otimes_R N$ and $N \otimes_R -$ are "right exact".

(v) If M is the cokernel of an injection $M_1 \hookrightarrow M_2$, then (ii) of the proposition does *not* let us write $(M_2/M_1) \otimes_R N$ as $(M_2 \otimes_R N)/(M_1 \otimes_R N)$, since $M_1 \otimes_R N \to M_2 \otimes_R N$ may not be injective. We just have an (explicit) isomorphism

$$(M_2/M_1) \otimes_R N \cong \mathrm{coker}(M_1 \otimes_R N \to M_2 \otimes_R N). \tag{28}$$

For an explicit example as to why, see Remark 6.13.

(vi) From the proof below, you can see that that both (i) and (ii) of the proposition can be immediately generalized to tensor products of $r > 2$ factors (but where all the variation happens at only one factor). For brevity, we will not state it formally.

(vii) For another (perhaps more usual, possibly more succinct/simpler to describe) proof of (ii) of the proposition, see Remark 6.19.

Proof of Proposition 6.10. For $j \in J$, let $\iota_j : M_j \hookrightarrow \bigoplus_i M_i$ be the obvious inclusion. For (i), we use the following diagram, which is functorial in L:

$$\begin{array}{ccc} \mathrm{Hom}_R(\bigoplus_i M_i \otimes_R N, L) & \xrightarrow{-\circ(((m_i)_i,n)\mapsto(m_i\otimes n)_i)} & \mathrm{Bil}_R(\bigoplus_i M_i, N; L) \\ {\scriptstyle (-\circ\iota_i\otimes\mathrm{id}_N)_i}\downarrow \cong & & \cong \downarrow {\scriptstyle (-\circ(\iota_i\times\mathrm{id}_N))_i} \\ \prod_i \mathrm{Hom}_R(M_i \otimes_R N, L) & \xrightarrow{(-\circ((m_i,n)\mapsto m_i\otimes n))_i} & \prod_i \mathrm{Bil}_R(M_i, N; L) \end{array} .$$

The vertical arrows are readily seen to be isomorphisms, and the bottom horizontal arrow is an isomorphism by the defining property of the $M_i \otimes_R N$. Therefore, the top horizontal arrow exists and is an isomorphism as well, and hence, by definition, gives a tensor product of $\bigoplus_i M_i$ and N. This implies that $(\bigoplus_i M_i) \otimes_R N$ exists and is isomorphic to $\bigoplus_i (M_i \otimes_R N)$. Moreover, check that the top horizontal arrow is given by $- \circ (((m_i)_i, n) \mapsto (m_i \otimes n)_i)$ as marked, so that the resulting isomorphism $(\bigoplus_i M_i) \otimes_R N \to \bigoplus_i (M_i \otimes_R N)$ indeed satisfies the requirement that $(m_i)_i \otimes n$ be sent by it to $(m_i \otimes n)_i$, as claimed.

The proof of (ii) is a straightforward adaptation of that of Proposition 6.6(ii), so we will be brief. Setting $\tilde{M} := \mathrm{coker}(f \otimes \mathrm{id} : M_1 \otimes_R N \to M_2 \otimes_R N)$, check that we have the following diagram which commutes, whose left square is analogous to the square in the

[10] Soon to be established as functors.

proof of Proposition 6.6(ii):

$$\begin{array}{ccccc} 0 \longrightarrow \mathrm{Bil}_R(M,N;L) & \longrightarrow & \mathrm{Bil}_R(M_2,N;L) & \longrightarrow & \mathrm{Bil}(M_1,N;L) \\ \exists \uparrow \cong & & \cong \uparrow -\circ(M_2\times N\to M_2\otimes_R N) & & \cong \uparrow -\circ(M_1\times N\to M_1\otimes_R N) \\ 0 \longrightarrow \mathrm{Hom}_R(\tilde{M},L) & \longrightarrow & \mathrm{Hom}_R(M_2\otimes_R N,L) & \longrightarrow & \mathrm{Hom}(M_1\otimes_R N;L) \end{array} .$$

Check that the rows of the above diagram are exact: the exactness of the bottom row is a simple fact seen in Lemma 6.18 below, while the exactness of the top row asserts that a bilinear map $M_2 \times N \to L$ restricts via $f \times \mathrm{id}_N$ to the zero bilinear map $M_1 \times N \to L$ if and only if it is a pullback, under $g \times \mathrm{id}_N$, of a bilinear form on $M \times N$.

But this means that the middle vertical map restricts to give an isomorphism $\mathrm{Hom}_R(\tilde{M}, L) \to \mathrm{Bil}_R(M, N; L)$, the dotted arrow of the above diagram. Hence, a tensor product $M \otimes_R N$ of M and N over R is given by the map $M \times N \to \tilde{M} \cong \mathrm{coker}(f \otimes \mathrm{id}_N)$ that, for $m = g(m_2) \in M$ and $n \in N$, sends (m, n) to the image of $m_2 \otimes n \in M_2 \otimes_R N$ in $\mathrm{coker}(f \otimes \mathrm{id}_N) = \tilde{M}$. [11]

Thus, we have an exact sequence

$$M_1 \otimes_R N \xrightarrow{f\otimes \mathrm{id}} M_2 \otimes_R N \xrightarrow{\tilde{g}} \tilde{M} = M \otimes_R N \to 0$$

(the exactness following from the definition of $\tilde{M}$ as $\mathrm{coker}(f \otimes \mathrm{id})$), where the map $M_2 \otimes_R N \to M \otimes_R N$ sends $m_2 \otimes n$ to $m \otimes n$ whenever $m = g(m_2)$ (this follows from the last sentence of the previous paragraph). In other words, the map $M_2 \otimes_R N \to M \otimes_R N$ above is exactly $g \otimes \mathrm{id}$, finishing the proof. □

Example 6.12. If M_1 is a free R-module with basis $\{e_i\}_i$ and M_2 is a free R-module with basis $\{f_j\}_j$, then by Proposition 6.10(i), it is easy to see that that $M_1 \otimes_R M_2$ is free with a basis given by $\{e_i \otimes f_j \mid i, j \in J\}$.

Let us discuss an example:

Remark 6.13. Rishiraj asked the timely question as to whether the tensor product is "exact" rather than just "right exact": i.e., if a sequence

$$0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M \to 0$$

of R-modules is exact, then (assuming all tensor products exist) is the sequence

$$0 \to M_1 \otimes_R N \xrightarrow{f\otimes \mathrm{id}_N} M_2 \otimes_R N \xrightarrow{g\otimes \mathrm{id}_N} M \otimes_R N \to 0$$

exact? The answer is no. By Proposition 6.10(ii), this question is equivalent to the following: if $f : M_1 \hookrightarrow M_2$ is injective, is $f \otimes \mathrm{id}_N : M_1 \otimes_R N \to M_2 \otimes_R N$ also injective? A counterexample is given by taking $f : M_1 \to M_2$ to be the map $\times 2 : \mathbb{Z} \to \mathbb{Z}$ given by multiplication by 2, and N to be $\mathbb{Z}/2\mathbb{Z}$. Then use Proposition 6.6 to see that the map

[11] See this by chasing the left square of the diagram.

$f \otimes \mathrm{id}_N : M_1 \otimes_R N \to M_2 \otimes_R N$ is the map $\times 2 : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$, which is the zero map, and hence not injective.

Later, we will study about *flat* R-modules M: those for which $- \otimes_R M$ is exact.

As remarked, the proof of Proposition 6.10(ii) is a generalization of that of Proposition 6.6(ii), so the latter can be seen from the former. Let $I \subset R$ be an ideal. Then by (28), for any R-module M, we have

$$M \otimes_R (R/I) \cong \mathrm{coker}(M \otimes_R I \to M \otimes_R R) \cong \mathrm{coker}(M \otimes_R I \to M) \cong M/IM,$$

since, although $M \otimes_R I \to M$ may not be injective, it takes $\sum m_j \otimes i_j$ to $\sum m_j \otimes i_j \in M \otimes R$, that is to say, to $\sum i_j m_j \in IM$, and therefore the image of $M \otimes_R I \to M$ is exactly IM.

Proposition 6.14. *For all R-modules $M_1, \dots, M_r$, a tensor product (M, u) of $M_1, \dots, M_r$ exists. Thus, sending $(M_1, \dots, M_r)$ to $\bigotimes_{i=1}^r M_i$ and $f_1, \dots, f_r$ to $\bigotimes_{i=1}^r f_i$ defines a functor*

$$\bigotimes_{i=1}^{r} : \underbrace{R\text{-}Mod \times R\text{-}Mod \times \cdots \times R\text{-}Mod}_{r \text{ factors}} \rightsquigarrow R\text{-}Mod.$$

Proof. Proof 1, when $r = 2$. Let us prove the first assertion (when $r = 2$). By Proposition 6.6, it is true when $M_1 = R$ or $M_2 = R$. Thus, by Proposition 6.10(i), it is true when M_1 or M_2 is free. Since any R-module is a cokernel of a map of free modules, Proposition 6.10(ii) implies that it, i.e., the first assertion, is true for general M_1 and M_2. The second assertion follows from the first, using Lemma 6.9. However, this involves the "strong axiom of choice": for each M_1, M_2, we are making a choice of $M_1 \times M_2 \to M_1 \otimes_R M_2$, which is a priori only well-defined up to a unique isomorphism.

This proof is easily generalized to the case of all $r \geqslant 2$ (see Remark 6.11(vi)), but (even for $r = 2$) it involves the strong axiom of choice. As Professor Nitsure alerted me to, the usual proof, given below, has the advantage of not using the strong axiom of choice.

Proof 2. Let $\tilde{M} = Free_R(M_1 \times \cdots \times M_r)$, the free R-module on the *set* $M_1 \times \cdots \times M_r$. Thus, pullback with respect to be obvious map $M_1 \times \cdots \times M_r \to \tilde{M}$, sending $(m_1, \dots, m_r)$ to $(m_1, \dots, m_r)$, induces an identification, for each set (and in particular R-module) L,

$$\mathrm{Hom}_R(\tilde{M}, L) = \mathrm{Mor}_{Set}(M_1 \times \cdots \times M_r, L).^{12} \tag{29}$$

Thus, our answer is going to be a *quotient* of $\tilde{M}$, since that is what it takes to cut down the collection of morphisms *from* $\tilde{M}$.

Let $N \subset \tilde{M}$ be the submodule generated by the elements of the following type:

$$(m_1, \dots, m_i + m_i', \dots, m_r) - (m_1, \dots, m_i, \dots, m_r) - (m_1, \dots, m_i', \dots, m_r),$$

$$(m_1, \dots, am_i, \dots, m_r) - a(m_1, \dots, m_r),$$

[12] Because we saw in Lecture 5 that $S \mapsto Free_R(S)$ is a left adjoint to the forgetful functor $R\text{-}Mod \rightsquigarrow Set$.

as the $1 \leqslant i \leqslant r$, the various $m_j \in M_j$ for $1 \leqslant j \leqslant r$, $m_i' \in M_i$ and $a \in R$ vary. Set $M := \tilde{M}/N$. Composing $M_1 \times \cdots \times M_r \to \tilde{M}$ with $\tilde{M} \to M$, we get a map $u : M_1 \times \cdots \times M_r \to M$.

It is immediate that u is linear in each variable, that is, it is multilinear.

Moreover, under the bijection (29), the subspace of $\mathrm{Hom}_R(\tilde{M}, L)$ that vanishes on N, namely $\mathrm{Hom}_R(M, L)$, maps to the collection of elements $f \in \mathrm{Mor}_{Set}(M_1 \times \cdots \times M_r, L)$ such that

$$f(m_1, \ldots, m_i + m_i', \ldots, m_r) = f(m_1, \ldots, m_i, \ldots, m_r) + f(m_1, \ldots, m_i', \ldots, m_r),$$

$$f(m_1, \ldots, am_i, \ldots, m_r) = af(m_1, \ldots, m_r),$$

as the $1 \leqslant i \leqslant r$, the various $m_j \in M_j$ for $1 \leqslant j \leqslant r$, $m_i' \in M_i$ and $a \in R$ vary. But this is, by definition, simply $\mathrm{Mult}_R(M_1, \ldots, M_r; L)$. In other words, (29) restricts to a bijection

$$\mathrm{Mult}_R(M_1, \ldots, M_r; L) \to \mathrm{Hom}_R(M, L).$$

It is clear that this bijection is induced by pullback with respect to u. The functoriality has already been taken care of, in Lemma 6.9. Here, no axiom of choice was involved, since the tensor product was explicitly constructed. □

Henceforth, we will not use the construction in Proof 2 of Proposition 6.14 at all.

6.2. **Hom-tensor adjointness.** Let M, N, L be R-modules. By the definition of bilinearity, we get bijections (in fact R-module isomorphisms)

$$\begin{array}{ccccc} \mathrm{Hom}_R(N, \mathrm{Hom}_R(M, L)) & \leftarrow \mathrm{Bil}_R(M, N; L) \to & \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, L)), \\ n \mapsto B(-, n) & \leftarrow\!\shortmid B \mapsto & (m \mapsto B(m, -)) \end{array} \tag{30}$$

that are functorial in M, N and L.

Thus, fixing M, we consider two functors $R\text{-}Mod \rightsquigarrow R\text{-}Mod$, given by $F = M \otimes_R -$ and $G = \mathrm{Hom}_R(M, -)$. We get a bijection:

$$\mathrm{Hom}_R(F(N), L) = \mathrm{Hom}_R(M \otimes_R N, L) \cong \mathrm{Bil}_R(M, N; L) \overset{(30)}{\to} \mathrm{Hom}_R(N, \mathrm{Hom}_R(M, L)) = \mathrm{Hom}_R(N, G(L)),$$

functorial in N and L, and hence realizing $M \otimes_R -$ as left adjoint to $\mathrm{Hom}_R(M, -)$. Similarly, using the other equality of (30), $- \otimes_R M$ is also left adjoint to $\mathrm{Hom}_R(M, -)$.

In other words, we have proved:

Proposition 6.15. *$M \otimes_R -$ and $- \otimes_R M$ are both left adjoint to $\mathrm{Hom}_R(M, -)$ (and hence in particular, by Lemma 5.2, naturally isomorphic to each other).*

We will redo the parenthetical natural isomorphism of $M \otimes_R -$ and $- \otimes_R M$ in Proposition 6.20 later below. Note that it agrees with earlier results such as Proposition 6.6.

Corollary 6.16. *$M \otimes_R -$ (or equivalently, $- \otimes_R M$) is cocontinuous, i.e., it commutes with small colimits.*

Proof. Since $M \otimes_R -$ and $- \otimes_R M$ are left adjoint to some functor by Proposition 6.15, they both preserve small colimits (Proposition 5.5). □

Remark 6.17. (i) Since coproducts are colimits, and so are cokernels ($\mathrm{coker}(f)$ is the coequalizer of f and 0), Corollary 6.16 gives a simultaneous proof of the parts (i) and (ii) of Proposition 6.10.

(ii) Conversely, Proposition 6.10 can yield a proof of Corollary 6.16 as well. Namely, one can show that, for categories having all small colimits (like R-Mod), a functor that preserves small coproducts and coequalizers also preserves small colimits (and a functor that preserves small products and equalizers preserves small limits, if the categories involved have all small limits). This shouldn't be surprising, given that we have seen how the existence of small products and equalizers (resp., small coproducts and coequalizers) implies the existence of small limits (resp., small colimits).

Now we note the analogue of Proposition 6.10(ii) for the functors $\mathrm{Hom}_R(N, -)$ and $\mathrm{Hom}_R(-, N)$: these are much easier than Proposition 6.10(ii).

Lemma 6.18. *(i) For an exact sequence $0 \to M \xrightarrow{g} M_1 \xrightarrow{f} M_2$ of R-modules,*

$$0 \to \mathrm{Hom}_R(N, M) \xrightarrow{g\circ -} \mathrm{Hom}_R(N, M_1) \xrightarrow{f\circ -} \mathrm{Hom}_R(N, M_2)$$

is exact. In other words, $\mathrm{Hom}_R(N, -) : R\text{-}Mod \rightsquigarrow R\text{-}Mod$ is "left exact".

(ii) For an exact sequence $M_1 \xrightarrow{f} M_2 \xrightarrow{g} M \to 0$ of R-modules,

$$0 \to \mathrm{Hom}_R(M, N) \xrightarrow{-\circ g} \mathrm{Hom}_R(M_2, N) \xrightarrow{-\circ f} \mathrm{Hom}_R(M_1, N)$$

is exact. In other words, $\mathrm{Hom}_R(-, N) : R\text{-}Mod^{op} \rightsquigarrow R\text{-}Mod$ is "left exact".

Proof. Proof 1. Both parts of the proposition are easy to verify directly. Namely, the first says that homomorphisms into M_1 which vanish on applying f land in $\ker f$ = image g, which means they factor through $M \xrightarrow{g} M_1$. The second says that homomorphisms from M_2 to N whose "restriction" (via f) to M_1 vanish, factor through $M_2/\mathrm{image}(f) = M_2/\ker g \cong M$.

Proof 2. The first part is a special case of the assertion that $\mathrm{Hom}_R(N, -)$ preserves limits in R-Mod, or equivalently, Set: but that is how limits in general categories $\mathcal{C}$ were defined, namely, to ensure that any $\mathrm{Mor}_{\mathcal{C}}(X, -)$ preserves limit. The second part is a special case of the assertion that $\mathrm{Hom}_R(-, N)$ takes colimits in R-Mod, or equivalently limits in R-Mod^{op}, to limits in R-Mod or equivalently, in Set: again, this is how colimits were defined.

Proof 3. (this is a bit too artificial though) $\mathrm{Hom}_R(N, -)$, being a left-adjoint (to $N \otimes_R -$), preserves small limits (Proposition 5.5). As for $\mathrm{Hom}_R(-, N)$: since

$$\mathrm{Hom}_R(L, \mathrm{Hom}_R(M, N)) \cong \mathrm{Bil}_R(L \times M, N) \cong \mathrm{Hom}_R(M, \mathrm{Hom}_R(L, N)) = \mathrm{Hom}_{R\text{-}Mod^{op}}(\mathrm{Hom}_R(L, N), M),$$

it follows that $\mathrm{Hom}_R(-, N)$, viewed as a functor R-$Mod^{op} \rightsquigarrow R$-$Mod$, is right adjoint to $\mathrm{Hom}_R(-, N)$, but this time viewed as a functor R-$Mod \rightsquigarrow R$-Mod^{op}, and is hence left exact. □

Remark 6.19. (i) A moral of some of the discussion above is that "left adjoint is right exact and right adjoint is left exact" (convince yourself of this; but this is informal since we will only formally define "right exact" and "left exact" later, when we (hopefully) do abelian categories).

(ii) While Lemma 6.18 tells us that $\mathrm{Hom}_R(N,-)$ and $\mathrm{Hom}_R(-,N)$ are left exact when appropriately interpreted, neither is exact: applying $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z},-)$ destroys the exactness of $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$, while applying $\mathrm{Hom}_{\mathbb{Z}}(-,\mathbb{Z})$ destroys the exactness of $0 \to \mathbb{Z} \stackrel{\times 2}{\to} \mathbb{Z}$ (Exercise: work out the details).

(iii) Later we will hopefully study about *projective* modules P ($\mathrm{Hom}_R(P,-)$ is exact), and *injective* modules I ($\mathrm{Hom}_R(-,I)$ is exact).

(iv) One way to articulate the part of the proof of Proposition 6.10(ii) that describes $M \otimes_R N$, is to use Lemma 6.18(ii), which gives (in the notation of Proposition 6.10(ii)) the exactness of

$$0 \to \mathrm{Hom}_R(M, \mathrm{Hom}_R(N,L)) \stackrel{-\circ g}{\to} \mathrm{Hom}_R(M_2, \mathrm{Hom}_R(N,L)) \stackrel{-\circ f}{\to} \mathrm{Hom}_R(M_1, \mathrm{Hom}_R(N,L)),$$

or equivalently that of

$$0 \to \mathrm{Hom}_R(M \otimes_R N, L) \stackrel{-\circ g\otimes \mathrm{id}_N}{\to} \mathrm{Hom}_R(M_2 \otimes_R N, L) \stackrel{-\circ f\otimes \mathrm{id}_N}{\to} \mathrm{Hom}_R(M_1 \otimes_R N, L).$$

This means precisely that $M_2 \otimes_R L \stackrel{g\otimes \mathrm{id}_N}{\to} M \otimes_R N$ is the coequalizer of $f \otimes \mathrm{id}_N$ and 0, or in other words, a cokernel of $f \otimes \mathrm{id}_N$, proving the exactness assertion in Proposition 6.10(ii).

(v) Much of what we have done today looks like a form of Proposition 5.5 (left adjoint functors preserve small colimits and right adjoint functors preserve small limits).

6.3. **Some other properties of tensor products.** You can check that the natural isomorphism in the proposition below coincides with that in Proposition 6.15.

Proposition 6.20. *For R-modules M and N, there is a unique isomorphism of R-modules $M \otimes_R N \to N \otimes_R M$ that maps each $m \otimes n$ to $n \otimes m$. The collection of these isomorphisms is natural in M and N, i.e., forms a natural isomorphism between functors R-Mod $\times$ R-Mod $\rightsquigarrow$ R-Mod (given by $(M,N) \rightsquigarrow M \otimes_R N$ and $(M,N) \rightsquigarrow N \otimes_R M$).*

Proof. We have an isomorphism of functors $\mathrm{Bil}_R(M,N;-) \to \mathrm{Bil}_R(N,M;-)$ obtained by "swapping M and N". This implies that

$$N \times M \stackrel{\text{swap}}{\to} M \times N \to M \otimes_R N$$

is a tensor product for $N \times M$, taking (n,m) to $m \otimes n$. By the uniqueness of tensor products, there is a unique isomorphism from $N \otimes_R M \to M \otimes_R N$, that transports $N \times M \to N \otimes_R M$ to $N \times M \to M \otimes_R N$, or in other words, takes $n \otimes m$ to $m \otimes n$ for each $m \in M, n \in N$. $\square$

Proposition 6.21. *If L, M, N are modules over R, there is are isomorphisms*

$$(L \otimes_R M) \otimes_R N \stackrel{a}{\to} L \otimes_R M \otimes_R N \stackrel{b}{\leftarrow} L \otimes_R (M \otimes_R N),$$

where either arrow is the unique isomorphism between its source and its target satisfying the property that for all $l \in L, m \in M$ and $n \in N$, the corresponding relation below holds:

$$(l \otimes m) \otimes n \overset{a}{\mapsto} l \otimes m \otimes n \overset{b}{\mapsfrom} l \otimes (m \otimes n).$$

.

Sketch of the proof of Proposition 6.21. We will prove the assertions involving a. For any R-module P, we have

(31)

$$\begin{aligned}\mathrm{Hom}_R((L \otimes_R M) \otimes_R N, P) &\cong \mathrm{Bil}_R((L \otimes_R M), N; P) \overset{(30)}{\cong} \mathrm{Hom}_R(L \otimes_R M, \mathrm{Hom}_R(N, P)) \\ &\cong \mathrm{Bil}_R(L, M; \mathrm{Hom}_R(N, P)) \cong \mathrm{Mult}_R(L, M, N; P) \cong \mathrm{Hom}_R(L \otimes_R M \otimes_R N, P),\end{aligned}$$

where the second-to-last isomorphism uses an argument analogous to that for (30): it sends $B : L \times M \to \mathrm{Hom}_R(N, P)$ to $A : (l, m, n) \mapsto B(l, m)(n)$. This being functorial in P, (31) gives a natural isomorphism between functors corepresented by $(L \otimes_R M) \otimes_R N$ and $L \otimes_R M \otimes_R N$, and hence also an isomorphism $(L \otimes_R M) \otimes_R N \to L \otimes_R M \otimes_R N$. It remains to prove that this isomorphism sends $(l \otimes m) \otimes n$ to $l \otimes m \otimes n$.

Chase through the isomorphisms in (31) and verify that, under that chain, if $\varphi \in \mathrm{Hom}_R((L \otimes_R M) \otimes_R N, P)$ corresponds to $A \in \mathrm{Mult}_R(L, M, N; P)$ and $\psi \in \mathrm{Hom}_R(L \otimes_R M \otimes_R N, P)$, then $\varphi((l \otimes m) \otimes n) = A(l, m, n) = \psi(l \otimes m \otimes n)$. This forces the above isomorphism $(L \otimes_R M) \otimes_R N \to L \otimes_R M \otimes_R N$ to send $(l \otimes m) \otimes n$ to $l \otimes m \otimes n$. □

Of course, the above lemma immediately generalizes to a sort of "associativity for arbitrarily bracketed tensor products", which we may use in what follows.

7. Lecture 7 – tensor products, the case of noncommutative rings

7.1. **Definition of tensor products over noncommutative rings.** If M, L are left modules over a noncommutative ring R, then $\mathrm{Hom}_R(M, L)$ is only an abelian group, and not an R-module: for $\varphi \in \mathrm{Hom}_R(M, L)$ and $r \in R$, $m \mapsto r\varphi(m)$ is no longer an R-module homomorphism, since $r \cdot \varphi(sm) = rs\varphi(m) \neq s \cdot r\varphi(m)$ in general.

Similarly, for left modules M, N over a noncommutative ring R, defining tensor products using R-bilinear maps $M \times N \to L$ would not work well, since that would cause a lot of collapsing: if $B : M \times N \to L$ is bilinear in the sense defined in Lecture 6, then

$$rsB(m, n) = rB(m, sn) = B(rm, sn) = sB(rm, n) = srB(m, n),$$

forcing each $rs - sr$ to annihilate the image of B.

Instead, the "correct" definition turns out to involve a right R-module with a left R-module:

Definition 7.1. Let $M \in \mathrm{Ob}\, Mod\text{-}R$ be a right R-module, and $N \in \mathrm{Ob}\, R\text{-}Mod$ a left R-module.

(i) For an abelian group L, a map $B : M \times N \to L$ is said to be R-middle linear (sometimes the word "balanced" is used instead) if it is biadditive (i.e., $\mathbb{Z}$-bilinear), and if for all $r \in R, m \in M$ and $n \in N$ we have:

$$B(mr, n) = B(m, rn).$$

(ii) Write $\mathrm{Midlin}_R(M, N; L)$ for the abelian group (under addition in L) of R-middle linear maps $M \times N \to L$, and $\mathrm{Midlin}_R(M, N; -) : AbGrp \rightsquigarrow AbGrp$ for the functor $L \rightsquigarrow \mathrm{Midlin}_R(M, N; L)$.

(iii) A tensor product of M and N over R is an abelian group $M \otimes_R N$, together with an R-middle linear map $u : M \times N \to M \otimes_R N$, such that for all abelian groups L, the following map is a bijection:

$$\mathrm{Hom}_{AbGrp}(M \otimes_R N, L) \overset{-\circ u}{\to} \mathrm{Midlin}_R(M, N; L).$$

Thus, by the Yoneda lemma, a tensor product $M \otimes_R N$ is of M and N over R an object of $AbGrp$ corepresenting the functor $\mathrm{Midlin}_R(M, N; -) : AbGrp \rightsquigarrow AbGrp$, together with an associated corepresentation.

Clearly, a tensor product of M and N is uniquely unique if it exists.

Proposition 7.2. *Let R be a commutative ring. Then $M \otimes_R N$, as defined above, exists and coincides with the abelian group underlying the Lecture 6 version of $M \otimes_R N$.*

You can try to prove this as an exercise; in any case, we will give a proof later in this lecture.

As in Lecture 6, if a tensor product of M and N over R exists, we will denote it and its underlying abelian group by $M \otimes_R N$, and the map $u : M \times N \to M \otimes_R N$ will be denoted by $(m, n) \mapsto m \otimes n$. As in Lecture 6 again, it is easy to see that every element of $M \otimes_R N$

is (highly non-uniquely) a $\mathbb{Z}$-linear combination of the $m_i \otimes n_i$, with each $m_i \in M$ and each $n_i \in N$.

Proposition 7.3. *Let R be a (not necessarily commutative, as usual) ring. Then for any right R-module M and a left R-module N, a tensor product $M \otimes_R N$ of M and N exists, and is functorial in M and N, i.e., this defines a functor*

$$Mod\text{-}R \times R\text{-}Mod \rightsquigarrow AbGrp.$$

Proof. One can adapt to this situation 'Proof 2' of the analogous assertion from Lecture 6. Namely, one considers the free abelian group (rather than the free R-module) on $M \times N$, and quotients it by relations of the form $(m+m', n)-(m, n)-(m', n)$, $(m, n+n')-(m, n)-(m, n')$ and $(mr, n) - (m, rn)$ as $m, m' \in M, n, n' \in N$ and $r \in R$ vary. □

7.2. R-S-bimodules.

Definition 7.4.

(i) An R-S-bimodule is an abelian group M given the structure of a left R-module and a right S-module such that the actions of R of S on M commute: $r \cdot (m \cdot s) = (r \cdot m) \cdot s$ for all $m \in M$, $r \in R$ and $s \in S$. In other words, we are given homomorphisms $R \to \mathrm{End}_{\mathbb{Z}}(M)$ and $S^{op} \to \mathrm{End}_{\mathbb{Z}}(M)$ with commuting images.

(ii) Write R-Mod-S for the category of R-S-bimodules (as usual, it is clear what R-S-bimodule homomorphisms should mean), and $\mathrm{Hom}_{R,S}(-,-)$ for morphisms in this category.

(iii) If M is an S-R-bimodule and N is an R-S'-bimodule, then $M \otimes_R N$ has a structure of an S-S'-bimodule, satisfying $s(m \otimes n)s' = sm \otimes ns'$ – prove this as an easy exercise using the universal property of the tensor product multiple times – and will be regarded as one.

(iv) For any S-R-bimodule L, we will denote by $\mathrm{Midlin}_{R,(S,S')}(M, N; L)$ the set of those $B \in \mathrm{Midlin}_R(M, N; L)$ that satisfy $B(sm, ns') = sB(m, n)s'$ for all $s \in S, s' \in S', m \in M$ and $n \in N$.

(v) If N is an R-S-bimodule, and L is an R-S'-module, $\mathrm{Hom}_R(N, L)$ is an (S, S')-bimodule under $(s \cdot f \cdot s')(n) = f(ns)s'$. [13] If L is simply a left R-module instead, then N and L can be regarded as $(R, \mathbb{Z})$-bimodules, so $\mathrm{Hom}_R(N, L)$ is then just a $(\mathbb{Z}, \mathbb{Z})$-bimodule, which in fact comes from just an abelian group.

Similarly, if N is an R-S'-bimodule and L is an S-S'-bimodule, then $\mathrm{Hom}_{S'}(N, L)$ is an S-R-bimodule via $(s \cdot f \cdot r)(n) = s \cdot f(rn)$.

Proposition 7.5. *If M is an S-R-bimodule and N is an R-S'-bimodule, the map $u : M \times N \to M \otimes_R N$ satisfies the following universal property: for any $L \in \mathrm{Ob}\, S$-Mod-S', the following is a well-defined bijection:*

$$\mathrm{Hom}_{S,S'}(M \otimes_R N, L) \overset{-\circ u}{\to} \mathrm{Midlin}_{R,(S,S')}(M, N; L).$$

[13] Note that, since S acts on N on the right, its action on maps f from N into any space by $(s \cdot f)(n) = f(ns)$ is a left action: when you take an action into an argument of a function, a left action becomes a right action and vice versa.

Proof. We already have the bijection between abelian groups that contain either side:

$$\mathrm{Hom}_{\mathbb{Z}}(M \otimes_R N, L) \stackrel{-\circ u}{\to} \mathrm{Midlin}_R(M, N; L).$$

u is clearly (S, S')-bilinear, so if $\varphi \in \mathrm{Hom}_{\mathbb{Z}}(M \otimes_R N, L)$ and $B = \varphi \circ u \in \mathrm{Midlin}_R(M, N; L)$, then

$$\begin{aligned} B \in \mathrm{Midlin}_{R,(S,S')}(M, N; L) &\iff B(sm, ns') = sB(m, n)s', \forall\, s \in S, m \in M, n \in N, s' \in S', \\ &\iff \varphi(sm \otimes ns') = s\varphi(m \otimes n)s', \forall\, s \in S, m \in M, n \in N, s' \in S', \\ &\iff \varphi \in \mathrm{Hom}_{S,S'}(M \otimes_R N, L), \end{aligned}$$

since the $m \otimes n$ span $M \otimes_R N$. □

Proof of Proposition 7.2. First we claim that $M \otimes_R N$, defined as above, has an obvious R-module structure.

For any commutative ring R, an R-module can be thought of as an R-R-bimodule, via $r \cdot m \cdot s = (r \cdot s) \cdot m$. In the converse direction, note that an R-R-bimodule L arises in this way from an R-module if and only if $a \cdot m = m \cdot a$ for all $a \in R$ and $m \in M$. Thus, the R-R-bimodule $M \otimes_R N$ arises this way from an R-module structure on $M \otimes_R N$: $a(m \otimes n) = am \otimes n = ma \otimes n = m \otimes an = m \otimes na = (m \otimes n)a$.

Then for any R-module L, viewing both $M \otimes_R N$ and L as R-R-bimodules, inside the sets $\mathrm{Hom}_{\mathbb{Z}}(M \otimes_R N, L)$ and $\mathrm{Midlin}_R(M \times N; L)$, the subsets $\mathrm{Hom}_{R,R}(M \otimes_R N, L)$ and $\mathrm{Midlin}_{R,(R,R)}(M \times N; L)$ coincide with $\mathrm{Hom}_R(M \otimes_R N, L)$ and $\mathrm{Bil}_R(M, N; L)$ respectively (check this). Therefore, by Proposition 7.5 (for the arrow involving "$-\circ u$" in the following sequence), we get bijections

$$\mathrm{Hom}_R(M \otimes_R N, L) = \mathrm{Hom}_{R,R}(M \otimes_R N, L) \stackrel{-\circ u}{\to} \mathrm{Midlin}_{R,(R,R)}(M \times N, L) = \mathrm{Bil}_R(M, N; L),$$

showing that $M \otimes_R N$, with its R-module structure as defined above, is a tensor product of M and N as defined in Lecture 6. □

Moral. Thus, unlike what Definition 7.1 might naively suggest, tensor product for noncommutative rings does satisfactorily generalize the tensor product for commutative rings: only, for a commutative ring R, an R-module M should be viewed as an (R, R)-bimodule when we look at it as per the theory for noncommutative rings.

Lemma 7.6. *If M_i is an R_{i-1}-R_i-bimodule for $i = 1, 2, 3$, then we have a unique isomorphism of R_0-R_3-bimodules:*

$$(M_1 \otimes_{R_1} M_2) \otimes_{R_2} M_3 \cong M_1 \otimes_{R_1} (M_2 \otimes_{R_2} M_3),$$

taking each $(m_1 \otimes m_2) \otimes m_3$ to $m_1 \otimes (m_2 \otimes m_3)$.

Sketch of proof. Check that both sides corepresent, in the category of R_0-R_3-bimodules, the functor that sends L to the abelian group of maps $\psi : M_1 \times M_2 \times M_3 \to L$ that are $\mathbb{Z}$-trilinear and satisfy $\psi(m_1 r_1, m_2, m_3) = \psi(m_1, r_1 m_2, m_3), \psi(m_1, m_2 r_2, m_3) = \psi(m_1, m_2, r_2 m_3)$, and $\psi(r_0 m_1, m_2, m_3 r_3) = r_0 \psi(m_1, m_2, m_3) r_3$ whenever $m_i \in M_i$ and $r_i \in R_i$ for $\in \{0, 1, 2, 3\}$. □

One way to view the lemma is as follows. Just like the binary tensor products above, we can define $M_1 \otimes_{R_1} \otimes M_2 \otimes \cdots \otimes_{R_{n-1}} M_n$, where M_1 is a right R_1-module, M_n is a left R_{n-1}-module, and for $2 \leqslant i \leqslant n-1$, each M_i is an (R_{i-1}, R_i)-bimodule. If further M_1 extends to an (R_0, R_1)-bimodule and M_n extends to an (R_{n-1}, R_n)-bimodule, then $M_1 \otimes_{R_1} \otimes M_2 \otimes_{R_2} \cdots \otimes_{R_{n-1}} M_n$ is an (R_0, R_n)-bimodule. All of these are proved as in the binary case. Then the above lemma can be justified by arguing that both modules agree with the unbracketed tensor product $M_1 \otimes_{R_1} M_2 \otimes_{R_2} M_3$, as an (R_0, R_3)-bimodule.

However, we will not deal much with nonbinary tensor products for noncommutative rings in the rest of this course.

7.3. **Hom-tensor adjointness, noncommutative case.** Now check as an exercise that for any (S, R)-bimodule M, (R, S')-bimodule N, and an (S, S')-bimodule L, we have bijections (in fact isomorphisms of abelian groups):

$$\mathrm{Hom}_{(R,S')}(N, \mathrm{Hom}_S(M, L)) \quad \leftarrow \mathrm{Midlin}_{R,(S,S')}(M, N; L) \rightarrow \quad \mathrm{Hom}_{(S,R)}(M, \mathrm{Hom}_{S'}(N, L)),$$

$$(n \mapsto B(-, n)) \leftarrow\!\shortmid B \mapsto \quad (m \mapsto B(m, -)).$$

Check also that the inverses are given by $\psi \mapsto ((m, n) \mapsto \psi(n)(m))$ and $\zeta \mapsto ((m, n) \mapsto \zeta(m)(n))$.

We have seen maps of this form earlier, from when we proved Hom-tensor adjointness in the commutative case. Therefore, we know that these work when $R = S = S' = \mathbb{Z}$, so what you need to check to do this exercise is that the conditions imposed by R, S and S' on either side are compatible: for example, if B is R-middle linear, then under $n \mapsto B(-, n)$, we have $rn \mapsto B(- \cdot r, n)$, which is $r \cdot B(-, n)$ by the definition of how $\mathrm{Hom}_S(M, L)$ is viewed as a left R-module.

Using the identification $\mathrm{Midlin}_{R,(S,S')}(M, N; L) \to \mathrm{Hom}_{(S,S')}(M \otimes_R N, L)$ of Proposition 7.5, whose inverse is $\varphi \mapsto ((m, n) \mapsto \varphi(m \otimes n))$, we get both the assertions of the following proposition:

Proposition 7.7. *(i) Let M be an S-R-bimodule. Then the functor $M \otimes_R - : R\text{-}Mod\text{-}S' \rightsquigarrow S\text{-}Mod\text{-}S'$ is left-adjoint to $\mathrm{Hom}_S(M, -) : S\text{-}Mod\text{-}S' \rightsquigarrow R\text{-}Mod\text{-}S'$: we have functorial bijections in $N \in \mathrm{Ob}\, R\text{-}Mod\text{-}S'$ and $L \in \mathrm{Ob}\, S\text{-}Mod\text{-}S'$:*

$$\mathrm{Hom}_{S,S'}(M \otimes_R N, L) \to \mathrm{Hom}_{R,S'}(N, \mathrm{Hom}_S(M, L)),$$

given by $\varphi \mapsto (n \mapsto (m \mapsto \varphi(m \otimes n)))$, with inverse taking $\psi \in \mathrm{Hom}_{R,S'}(N, \mathrm{Hom}_S(M, L))$ to an element of $\mathrm{Hom}_{S,S'}(M \otimes_R N, L)$ that maps each $m \otimes n$ to $\psi(n)(m)$.

(ii) Similarly, if N is is an R-S'-bimodule, then $- \otimes_R N : S\text{-}Mod\text{-}R \rightsquigarrow S\text{-}Mod\text{-}S'$ is left-adjoint to $\mathrm{Hom}_{S'}(N, -) : S\text{-}Mod\text{-}S' \rightsquigarrow S\text{-}Mod\text{-}R$. [14]

[14] Here is one way to remember this: for an R-S'-bimodule, the adjoint of tensoring with respect to R should be homming with respect to the other ring, namely S': the two functors should go in opposite direction, so should use up actions of different rings.

Corollary 7.8. *Let M be an S-R-bimodule. Then the functor $M \otimes_R - : R\text{-}Mod \rightsquigarrow S\text{-}Mod$ and $- \otimes_S M : Mod\text{-}S \rightsquigarrow Mod\text{-}R$ are right exact, while the functor $\mathrm{Hom}_S(M, -) : S\text{-}Mod \rightsquigarrow R\text{-}Mod$ is left exact.*

Proof. This follows from the fact that $M \otimes_R -$ and $- \otimes_R M$, being left adjoints, preserve small colimits and hence cokernels, and $\mathrm{Hom}_S(M, -)$, being a right adjoint, preserves small limits and hence kernels. [15] However, one can also prove this directly (especially, the latter is easy), using arguments from Lecture 6. □

Exercise 7.9. Recall the notion of units and counits of an adjunction, either from the notes for Lecture 4 or from HW 3. Show that, the adjunction of Proposition 7.7 between $M \otimes_R -$ and $\mathrm{Hom}_S(M, -)$ has unit and counit equal to the natural transformations

$$\mathrm{id} \to \mathrm{Hom}_S(M, M \otimes_R -) \qquad \text{and } M \otimes_R (\mathrm{Hom}_S(M, -)) \to \mathrm{id},$$

given by the maps $N \to \mathrm{Hom}_S(M, M \otimes_R N)$ and $M \otimes_R \mathrm{Hom}_S(M, L) \to L$, given respectively by $n \mapsto (m \mapsto m \otimes n)$ and the evaluation map $m \otimes \omega \mapsto \omega(m)$. Mentally relate these formulas to the prescriptions in Proposition 7.7 by staring long enough (the point being that the adjunctions are described explicitly using the unit and counit).

7.4. **Extension, restriction and coextension of scalars.** Let $R \to S$ be a homomorphism of not necessarily commutative rings. We would like to look at the special cases of Proposition 7.7 where $M = S$. Then M can be viewed either as an S-R-bimodule or as an R-S-bimodule, giving us two functors $R\text{-}Mod \rightsquigarrow S\text{-}Mod$, and two functors $S\text{-}Mod \rightsquigarrow R\text{-}Mod$:

Definition 7.10. (i) Viewing $M = S$ as an (S, R)-bimodule:

- $M \otimes_R -$ becomes the functor $S \otimes_R - : R\text{-}Mod \rightsquigarrow S\text{-}Mod$. This functor is called the *extension of scalars along $R \to S$*, arguably the most 'obvious' way of producing an S-module from an R-module.
- $\mathrm{Hom}_S(M, -) = \mathrm{Hom}_S(S, -) : S\text{-}Mod \rightsquigarrow R\text{-}Mod$. We have a functorial isomorphism $\mathrm{Hom}_S(S, L) \to L$, given by $\varphi \mapsto \varphi(1)$. Under this isomorphism, the left R-module structure on $\mathrm{Hom}_S(S, L)$ obtained from the S-R-bimodule structure on S transfers to the left R-module structure on L obtained as follows: L is already a left S-module, so R acts on it via $R \to S$. Thus, this functor is described more simply as: "given $N \in \mathrm{Ob}\, S\text{-}Mod$, take it to N itself, but viewed as in $\mathrm{Ob}\, R\text{-}Mod$, through $R \to S$". This functor is called the *restriction of scalars along $R \to S$.*"

(ii) Viewing M as an (R, S)-bimodule:

- This time, we have an isomorphism $M \otimes_S N = S \otimes_S N \cong N$ for each S-module N, and under this isomorphism, the R-module structure on $S \otimes_S N$, which comes from the left-multiplication by R on S, clearly transfers to the action of R on the S-module N via $R \to S$. Thus, $S \otimes_S - : S\text{-}Mod \rightsquigarrow R\text{-}Mod$ is the same as the restriction of scalars along $R \to S$ seen above.

[15] In general (for "additive functors"), left adjoint functors are right exact, and right adjoint functors are left exact.

- On the other hand, $\mathrm{Hom}_R(S,-) : R\text{-}Mod \rightsquigarrow S\text{-}Mod$ is called the *coextension of scalars along* $R \rightsquigarrow S$.

Thus, Proposition 7.7 immediately gives, on specializing to the above two situations:

Proposition 7.11. *Let $R \to S$ be a homomorphism of not necessarily commutative rings (but with 1, as always):*

(i) The extension of scalars functor $S \otimes_R - : R\text{-}Mod \rightsquigarrow S\text{-}Mod$ is left adjoint to the restriction of scalars functor $S\text{-}Mod \rightsquigarrow R\text{-}Mod$. More precisely, for any left R-module N, and left S-module L, we have a functorial isomorphism of abelian groups:

$$\mathrm{Hom}_S(S \otimes_R N, L) \overset{-\circ(n\mapsto 1\otimes n)}{\to} \mathrm{Hom}_R(N, L),$$

given by precomposition with the map $\iota \in \mathrm{Hom}_R(N, S \otimes_R N)$ given by $n \mapsto 1 \otimes n$ (on the right-hand side, L stands for L viewed as an R-module via $R \to S$).

(ii) The coextension of scalars functor $\mathrm{Hom}_R(S,-) : R\text{-}Mod \rightsquigarrow S\text{-}Mod$ is right adjoint to the restriction of scalars functor $S\text{-}Mod \rightsquigarrow R\text{-}Mod$. More precisely, for any left S-module N and left R-module L, we have a functorial isomorphism

$$\mathrm{Hom}_R(N, L) \to \mathrm{Hom}_S(N, \mathrm{Hom}_R(S, L)),$$

whose inverse given by post-composition with the map $\mathrm{Hom}_R(S, L) \to L$ defined by $\varphi \mapsto \varphi(1)$ (the map itself takes $\psi \in \mathrm{Hom}_R(N, L)$ to $n \mapsto (s \mapsto \psi(sn))$).

Proof. The above discussion tells us how to apply Proposition 7.7. More precisely:

- The prescription for the former isomorphism maps $\mathrm{Hom}_S(S \otimes_R N, L) \to \mathrm{Hom}_R(N, \mathrm{Hom}_S(S, L))$ by $\varphi \mapsto (n \mapsto (s \mapsto \varphi(s \otimes n)))$, which, upon identifying $\mathrm{Hom}_S(S, L)$ with L via $\psi \mapsto \psi(1)$, becomes $\varphi \mapsto (n \mapsto \varphi(1 \otimes n))$, which is as claimed.
- The prescription for the inverse of latter isomorphism maps (recalling that R and S are now interchanged) $\mathrm{Hom}_S(N, \mathrm{Hom}_R(S, L))$ to $\mathrm{Hom}_R(S \otimes_S N, L)$ by sending $\psi \in \mathrm{Hom}_S(N, \mathrm{Hom}_R(S, L))$ to the element of $\mathrm{Hom}_R(S \otimes_S N = N, L)$ that sends $1 \otimes n = n$ to $\psi(n)(1)$: in other words, this inverse is given by composing ψ with the map $\mathrm{Hom}_R(S, L) \to L$ taking ζ to $\zeta(1)$.

□

Example 7.12. (i) Let $R = \mathbb{R} \hookrightarrow \mathbb{C} = S$. Then Proposition 7.11(i) tells us that $\mathbb{R}$-linear maps from a real vector space V to a complex vector space W are in bijection with $\mathbb{C}$-linear maps from the *complexification* $V_{\mathbb{C}} := V \otimes_{\mathbb{R}} \mathbb{C}$ of V to W. Prove this elementarily: $V \subset V_{\mathbb{C}}$ has $V_{\mathbb{C}} = V \oplus iV$ as its $\mathbb{C}$-span, and any $\mathbb{R}$-linear map $V \to W$ extends uniquely to a $\mathbb{C}$-linear map $V_{\mathbb{C}} \to W$, and any $\mathbb{C}$-linear map $V_{\mathbb{C}} \to W$ arises this way. You might have seen this in many contexts.

(ii) Again, let $R = \mathbb{R} \hookrightarrow \mathbb{C} = S$. Proposition 7.11(ii) tells us that $\mathbb{R}$-linear maps from a complex vector space V to a real vector space W are in bijection with $\mathbb{C}$-linear

maps from V to the complex vector space $\mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, W)$ (which gets the $\mathbb{C}$-vector space structure from viewing the $\mathbb{C}$ in $\mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, W)$ as a right $\mathbb{C}$-module).

One 'popular' situation where this is applied is the deduction of the Hahn-Banach theorem in the complex case from the real case. Consider the particular case where $W = \mathbb{R}$, so we get a bijection $\mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R}) \to \mathrm{Hom}_{\mathbb{C}}(V, \mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, \mathbb{R}))$. But we can identify $\mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, \mathbb{R})$ with $\mathbb{C}$ as a $\mathbb{C}$-vector space, using the unique $\mathbb{C}$-linear isomorphism $\mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, \mathbb{R}) \to \mathbb{C}$ whose inverse takes $1 \in \mathbb{C}$ to the element "read off the real part" $\in \mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, \mathbb{R})$. Now you can check that the inverse of the bijection $\mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R}) \to \mathrm{Hom}_{\mathbb{C}}(V, \mathrm{Hom}_{\mathbb{R}}(\mathbb{C}, \mathbb{R})) = \mathrm{Hom}_{\mathbb{C}}(V, \mathbb{C})$ takes a linear functional $F \in \mathrm{Hom}_{\mathbb{C}}(V, \mathbb{C})$ to $\mathrm{Re}\, F \in \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$, which is what is used in the Hahn-Banach theorem for complex Banach spaces.

(iii) If R is a commutative ring and $I \subset R$ is an ideal, we have an "inclusion functor" from the category of R/I-modules to the category of R-modules, which is an equivalence onto the full subcategory of R-modules annihilated by I. Note that this functor is just the restriction of scalars functor along $R \to S := R/I$. In this case, you already know that a left-adjoint to this functor is given by $M \rightsquigarrow M/IM$, which is just saying that any morphism from M to a module N annihilated by I factors uniquely through $M \to M/IM$. This is a manifestation of Proposition 7.11(i), since according to it a left-adjoint is given by $R/I \otimes_R M$, and we have already seen that $R/I \otimes_R M \cong M/IM$. Thus, this basic fact about annihilators is included in the adjunction between the extension of scalars and the restriction of scalars.

(iv) Let R be an integral domain, and let K be its quotient field. The 'inclusion functor' from K-vector spaces to R-modules is restriction of scalars along $R \hookrightarrow K$, and has left-adjoint given by $- \otimes_R K$, from R-modules into K-vector spaces: any R-module map from an R-module M into a K-vector space uniquely factors through $M \to M \otimes_R K \cong (M/M_{tors}) \otimes_R K$.

If R is a principal ideal domain and a module M over R is isomorphic to $R^n \oplus (\bigoplus_i R/(d_i))$ with each d_i a nonunit, then $M \otimes_R K \cong K^n$. Thus, the uniqueness of the rank n of the "free" part of M can also be read off by tensoring with K. Similarly, if we write M as $R^n \oplus (\bigoplus_i R/(p_i^{a_i}))$, with p ranging over primes, the uniqueness of the contribution of the parts corresponding to a single p can be obtained by tensoring M with $R/(p^n)$ for a large enough n. The language of tensoring allows us to make the proof of the uniqueness assertion in the structure theorem for modules over a PID slightly simpler to write down, as a convenient linguistic tool, but in many other situations is tremendously simplifying and useful in arguably much more fundamental ways.

7.5. **Application to representation theory.** In this subsection, we will let k be any commutative ring (not necessarily a field). Given a group G, we will perhaps abusively refer to k-modules on which G acts as representations of G on k-modules: these are actions $G \times M \to M$ of G on k-modules M by k-module automorphisms, which can also be written $G \to \mathrm{Aut}_k(M)$. Denote their category by $Rep_k(G)$. However, despite this terminology, the results that follow work for general commutative rings k, and in the case where $k = \mathbb{Z}$, the

induced and coinduced representations that we talk of below coincide with induced and coinduced modules that one sees in group cohomology (where one studies groups acting on abelian groups, i.e., on $\mathbb{Z}$-modules).

As we now briefly recall, this category is isomorphic to the category of $k[G]$-modules, where $k[G]$ is the so called group algebra of G, a ring whose underlying set is the set $\sum_{g\in G} a_g g$ of formal k-linear combinations of elements of G, where $a_g = 0$ for all but finitely many elements of G, and multiplication is given by convolution:

$$\Big(\sum_g a_g g\Big)\Big(\sum_g b_g g\Big) = \sum_g \Big(\sum_{\substack{h,k\in G\\ hk=g}} a_h b_k\Big) g.$$

The multiplicative identity is $\sum a_g g$, where $a_g = 0$ unless $g = e$, the identity element of G, and where $a_e = 1$.

Specifically, we have an inclusion $G \hookrightarrow k[G]$, sending g to $\sum \delta_{g,g'} g'$, where $\delta_{g,g'}$ equals 1 or 0 depending on whether or not $g = g'$. Clearly, $G \subset k[G]$ is a basis, so any representation $\rho : G \to \mathrm{Aut}_k(M)$ extends k-linearly to a k-module homomorphism $k[G] \to \mathrm{End}_k(M)$, with $\sum_g a_g g$ acting by $\sum_g a_g \rho(g)$. Using the above definition of multiplication, it is easy to check this extension $k[G] \to \mathrm{End}_k(M)$ is a k-algebra homomorphism, and thus realizes M as a $k[G]$-module. Conversely, given a $k[G]$-module M given by $k[G] \to \mathrm{End}_k(M)$, we obtain a representation $G \to \mathrm{Aut}_R(M)$ by restricting to $G \subset k[G]^\times \subset k[G]$ (invertible elements in $R^\times$ act as automorphisms on any R-module).

Check that this allows us to identify $Rep_k(G)$ with $k[G]$-Mod.

Remark 7.13. Part of what we showed up above is that $G \rightsquigarrow k[G]$ is a left-adjoint to the functor from k-algebras to groups that sends R to $R^\times$. Note also that $G \rightsquigarrow k[G]$ is analogous to $S \rightsquigarrow Free_k(S)$, except that we have a group G here, and correspondingly a ring structure on $Free_k(S)$.

In what follows, we will use the identification between $Rep_k(G)$ and $k[G]$-Mod above the remark, to identify representations of any group G with $k[G]$-modules. If $\pi : G \to \mathrm{Aut}_k(M)$ is a representation of G, we will refer to M as the 'space of π', and refer to the representation also as (π, M).

Now let G be a group, and $H \subset G$ a subgroup. The inclusion $H \hookrightarrow G$ extends k-linearly to an inclusion $R := k[H] \hookrightarrow k[G] =: S$, which is a homomorphism of rings.

Our first task is to just translate Proposition 7.11 in this situation, from the language of R-modules and S-modules to the language of representations of H and G.

Definition 7.14. (i) The extension of scalars functor

$$k[G] \otimes_{k[H]} - : Rep_k(H) = k[H]\text{-}Mod \rightsquigarrow k[G]\text{-}Mod \rightsquigarrow Rep_k(G)$$

will also be referred to as induction of representations from H to G, and written $\pi \rightsquigarrow \mathrm{Ind}_H^G \pi$. $\mathrm{Ind}_H^G \pi$ is called the representation obtained by inducing π from H to G.

(ii) Similarly, the coextension of scalars functor $Rep_k(H) \to Rep_k(G)$, given by $\mathrm{Hom}_{k[H]}(k[G], -)$, is called coinduction of representations from H to G, and written $\pi \rightsquigarrow \mathrm{coInd}_H^G \pi$.

(iii) Similarly, the functor of restriction of representations from G to H – which amounts to viewing a $k[G]$-module as a $k[H]$ module via $k[H] \hookrightarrow k[G]$, denoted $\sigma \rightsquigarrow \sigma|_H$ or $\pi \rightsquigarrow Res_H^G \sigma$.

Remark 7.15. (i) Thus, by Proposition 7.11, $\mathrm{Ind}_H^G \pi$ comes with a map $\pi \hookrightarrow \mathrm{Ind}_H^G \pi$, composition with which induces a bijection, for any representation σ of G,

$$\mathrm{Hom}_G(\mathrm{Ind}_H^G \pi, \sigma) \to \mathrm{Hom}_H(\pi, \sigma|_H).$$

Moreover, it is immediate to check that for any representation (π, M) of H in $Rep_k(H)$, the description of $\rho := \mathrm{Ind}_H^G \pi$ as $k[G] \otimes_{k[H]} M$, agrees with the following description from Serre's book on linear representations of finite groups: ρ is obtained by inducing π from H to G if $\rho|_H$ contains a copy of π, whose translates by representatives for G/H direct sum to π (see Definition 3.3 of the book).

(ii) Similarly, $\mathrm{coInd}_H^G \pi$ comes with a map $\mathrm{coInd}_H^G \pi \to \pi$, composition with which induces a bijection, for any representation σ of G,

$$\mathrm{Hom}_G(\sigma, \mathrm{coInd}_H^G \pi) \to \mathrm{Hom}_H(\sigma, \pi|_H).$$

Moreover, we can describe coinduction at the level of representations as follows. Since $H \subset k[H]$ and $G \subset k[G]$ are free k-module bases, any element of (the underlying space of) $\mathrm{Hom}_{k[H]}(k[G], (\pi, M))$ is determined by its restriction to $G \subset k[G]$, which is a function $f : G \to M$, satisfying $f(hg) = \pi(h)f(g)$ for all $h \in G$. Thus, the elements of $\mathrm{coInd}_H^G \pi$ can be described as

$$\{f : G \to M \mid f(hg) = \pi(h)f(g) \,\forall\, h \in H, g \in G\}, \tag{32}$$

on which the action of G is given by right-multiplication: $(g \cdot f)(g') = f(g'g)$ (since the left $k[G]$-module structure on $\mathrm{Hom}_{k[H]}(k[G], (\pi, M))$ is given by right-multiplication by $k[G]$ on the argument).

(iii) On the other hand, the restriction of scalars functor $Rep_k(G) \rightsquigarrow Rep_k(H)$ is much easier: it is just viewing a representation σ of G on M as a representation of H on M by restricting the action of G to H.

(iv) All the three functors, $\mathrm{Ind}_H^G, \mathrm{coInd}_H^G$ and Res_H^G, are exact, as we now explain. For general rings $R \to S$, restriction of scalars along it is obviously exact, but extension and coextension of scalars may not be, though note that they are if S is free as an R-module. This is clearly the case when $R = k[H] \hookrightarrow k[G] = S$, with $H \subset G$.

Proposition 7.16. *If $H \subset G$ is of finite index, then the functors Ind_H^G and coInd_H^G of induction and coinduction are naturally isomorphic to each other.*

Proof. Proof 1. Let (π, M) be a representation of H. Define $\mathrm{coInd}_H^G(\pi, M) \to \mathrm{Ind}_H^G(\pi, M)$ by sending $(f : G \to M) \in \mathrm{coInd}_H^G(\pi, M)$ (realized as in (32)) to $\sum_{g \in [H\backslash G]} g^{-1} \otimes f(g) \in k[G] \otimes_{k[H]} M = \mathrm{Ind}_H^G(\pi, M)$, where $[H\backslash G]$ is a set of representatives for $H\backslash G$ in G: this sum is independent of the choice of representatives $[H\backslash G]$, since $(hg)^{-1} \otimes f(hg) =$

$g^{-1}h^{-1} \otimes h \cdot f(g) = g^{-1} \otimes f(g)$ (as the tensor product is taken over $k[H]$). This map is readily verified to be an isomorphism of k-modules, with a two-sided inverse sending $g \otimes m$ to the map $G \to M$ that sends any hg^{-1} with $h \in H$ to hm, and every element of $G\backslash(Hg^{-1})$ to 0. It also respects the G-action, since it takes $(g_0 f) : g \mapsto f(gg_0)$ to

$$\sum_{g \in [H\backslash G]} g^{-1} \otimes f(gg_0) = g_0 \cdot \sum_{g \in [H\backslash G]} (gg_0)^{-1} \sum_{g \in [H\backslash G]} (gg_0^{-1}) \otimes f(gg_0) = g_0 \cdot \sum_{g \in [H\backslash G]} g^{-1} \otimes f(g),$$

since $\{gg_0 \mid g \in [H\backslash G]\}$ is also a set of representatives for $H\backslash G$. Since this isomorphism is functorial in (π, M), it gives a natural isomorphism from coInd_H^G to Ind_H^G.

Proof 2. If S is a free left R-module of finite rank, then for any left R-module M, applying $- \otimes_R M$ gives an isomorphism $\mathrm{Hom}_R(S, R) \otimes_R M \to \mathrm{Hom}_R(S, M)$ of abelian groups (where $\mathrm{Hom}_R(S, R)$ is viewed as a right-R-module using the right R-module structure on R): indeed, this is clear if $S = R$, and the general case follows by taking direct sums (here the finite rank condition is necessary since $\mathrm{Hom}_R(-, R)$ converts a direct sum into a direct product).

If further $S \to \mathrm{Hom}_R(S, R)$ is an isomorphism of (S, R)-bimodules, then $S \otimes_R M \to \mathrm{Hom}_R(S, R) \otimes_R M$ is an isomorphism of left S-modules. Thus, if $R \to S$ is a ring homomorphism that makes S into a finite rank free right R-module, then to get a natural isomorphism $S \otimes_R - \to \mathrm{Hom}_R(S, -)$ of functors valued in S-modules, it is enough to get an isomorphism $S \to \mathrm{Hom}_R(S, R)$ of S-R-bimodules.

We will show this for $R = k[H] \hookrightarrow k[G] = S$. It is immediate that S is a free left R-module with basis given by a set $[H\backslash G]$ of representatives for $H\backslash G$, and similarly a free right R-module as well.

To get an isomorphism $S \to \mathrm{Hom}_R(S, R)$ of S-R-bimodules, it is enough to get a *perfect* pairing

$$(\cdot, \cdot) : S \times S \to R$$

such that $(ss_1, s_2) = (s_1, s_2 s)$ and $(s_1 r, s_2) = (s_1, s_2)r$, for all $s, s_1, s_2 \in S$ and $r \in R$: given such a pairing, $s \mapsto (s, -) \in \mathrm{Hom}_R(S, R)$ will give the desired isomorphism. Here, "perfect pairing" means a pairing such that $s \mapsto (s, -)$ defines an isomorphism $S \to \mathrm{Hom}_R(S, R)$ (so the previous sentence is a tautology, except you should check that the various actions match).

Define a "trace" $tr : k[G] \to k[H]$ by sending $\sum_{g \in G} a_g g$ to $\sum_{h \in H} a_h H$. Now define

$$(s_1, s_2) = tr(s_2 s_1).$$

The property $(ss_1, s_2) = (s_1, s_2 s)$ is immediate, while the property $(s_1 r, s_2) = (s_1, s_2)r$ follows from the fact that tr is a right-R-module homomorphism (even an (R, R)-bimodule homomorphism).

It remains to show that this pairing is perfect. For this, it is enough to show that there is a basis for (the first copy of) S as a free left R-module, that is dual for the pairing to some basis for (the second copy of) S as a free right R-module. Indeed, $\{g \mid g \in [H\backslash G]\}$ and $\{g^{-1} \mid g \in [H\backslash G]\}$ are such bases. $\square$

8. Lecture 8 – tensor products of algebras, and tensor algebras (Incomplete/extra crude)

8.1. **Tensor product of algebras.** Throughout today's lecture, R will denote a commutative ring. But other rings, like S, will not be assumed to be commutative. Recall that an R-algebra is a homomorphism of rings $\iota : R \to S$, such that $\iota(R)$ is contained in the center of S. Note that it is both unnecessary and "lossy" to require ι to be injective.

Let $j_1 : R \to S_1$ and $j_2 : R \to S_2$ be R-algebras. In this subsection, we would like to realize $S_1 \otimes_R S_2$ as an R-algebra.

Consider

$$S_1 \times S_2 \times S_1 \times S_2 \to S_1 \otimes_R S_2,$$

given by $(s_1, s_2, s_1', s_2') \mapsto s_1 s_1' \otimes s_2 s_2'$.

This map is clearly R-multilinear, and hence factors through a map

$$(S_1 \otimes_R S_2) \otimes_R (S_1 \otimes_R S_2) \cong S_1 \otimes_R S_2 \otimes_R S_1 \otimes_R S_2 \to S_1 \otimes_R S_2,$$

satisfying:

$$\Big(\sum_i s_{1,i} \otimes s_{2,i}\Big)\Big(\sum_j s_{1,j}' \otimes s_{2,j}'\Big) = \sum_{i,j} s_{1,i} s_{1,j}' \otimes s_{2,i} s_{2,i}'.$$

It is easy to check that this map, or 'operation', is associative, has multiplicative identity $1 \otimes 1$, and that it distributes over addition, making $S_1 \otimes_R S_2$ into a ring. Moreover, $r \mapsto j_1(r) \otimes 1 = 1 \otimes j_2(r)$ defines a ring homomorphism $R \to S_1 \otimes_R S_2$, with image in the center of $S_1 \otimes_R S_2$, so that $S_1 \otimes_R S_2$ is an R-algebra. If S_1 and S_2 are commutative, so is $S_1 \otimes_R S_2$.

It is clear that there are ring homomorphisms (in fact, R-algebra homomorphisms) $\iota_1 : S_1 \to S_1 \otimes_R S_2$ and $\iota_2 : S_2 \to S_1 \otimes_R S_2$, given by $\iota_1(s_1) = s_1 \otimes 1$ and $\iota_2(s_2) = 1 \otimes s_2$.

Remark 8.1. If S_1 is commutative, then $\iota_1 : S_1 \to S_1 \otimes_R S_2$ makes $S_1 \otimes_R S_2$ into an S_1-algebra, upgrading the R-algebra structure on it. Similarly with ι_2.

Proposition 8.2. *The tensor product is a coproduct in the category of commutative R-algebras: if $R \to S_1$ and $R \to S_2$ are commutative R-algebras, a coproduct of theirs is given by $(S_1 \otimes_R S_2, \iota_1, \iota_2)$, where $\iota_1 : S_1 \to S_1 \otimes_R S_2$ and $\iota_2 : S_2 \to S_1 \otimes_R S_2$ are as above.*

Remark 8.3. $S_1 \otimes_R S_2$ is *not* a coproduct of S_1 and S_2 in the category of not-necessarily-commutative R-algebras, even if S_1 and S_2 are themselves commutative: coproducts in the category of not-necessarily-commutative R-algebras, like coproducts in the category of nonabelian groups, are more complicated.

Proof of Proposition 8.2. For a commutative R-algebra S, what we need to prove is that the following map is a bijection:

$$\mathrm{Hom}_{R\text{-alg}}(S_1 \otimes_R S_2, S) \overset{(-\circ\iota_1, -\circ\iota_2)}{\to} \mathrm{Hom}_{R\text{-alg}}(S_1, S) \times \mathrm{Hom}_{R\text{-alg}}(S_2, S).$$

It is an injection because $\iota_1(S_1)$ and $\iota_2(S_2)$ generate $S_1 \otimes_R S_2$ as a ring, so let us see its surjectivity. Given $\varphi_1 \in \mathrm{Hom}_{R\text{-alg}}(S_1, S)$ and $\varphi_2 \in \mathrm{Hom}_{R\text{-alg}}(S_2, S)$, the map $S_1 \times S_2 \to S$ given by $(s_1, s_2) \mapsto \varphi_1(s_1)\varphi_2(s_2)$ is R-bilinear, and hence gives us an R-linear map $\varphi : S_1 \otimes_R S_2 \to S$ satisfying $\varphi(s_1 \otimes s_2) = \varphi_1(s_1)\varphi_2(s_2)$. φ respects multiplication at the level of 'pure tensors' $s_1 \otimes s_2$ and $s_1' \otimes s_2'$:

$$\varphi((s_1 \otimes s_2)(s_1' \otimes s_2')) = \varphi(s_1 s_1' \otimes s_2 s_2') = \varphi_1(s_1 s_1')\varphi_2(s_2 s_2') = \varphi_1(s_1)\varphi_2(s_2)\varphi_1(s_1')\varphi_2(s_2') = \varphi(s_1 \otimes s_2)\varphi(s_1' \otimes s_2'),$$

where the second step from the last uses that S is commutative. Therefore, by distributivity, φ respects multiplication in general. $\varphi(1 \otimes 1) = \varphi_1(1)\varphi_2(1) = 1$, so φ is a ring homomorphism. It is clear that $\varphi : S_1 \otimes_R S_2 \to S$ respects the R-algebra structures on the two sides, so it is an R-algebra homomorphism. Clearly, $\varphi \circ \iota_1 = \varphi_1$ and $\varphi \circ \iota_2 = \varphi_2$, proving the surjectivity. □

Remark 8.4. Recall that, purely terminologically, we defined the category of affine algebraic schemes over k to be the category opposite to that of commutative finitely generated k-algebras. Therefore, the product in this category is given by the coproduct in the category of finitely generated commutative k-algebras, which by Proposition 8.2 is given by the tensor product of k-algebras. For those of you who have seen some basic algebraic geometry, this (plus some "reducedness/irreducibility considerations") is why a product of affine varieties is described by taking the tensor product of the coordinate rings. Here is how this 'productness' is realized at the level of 'k-points' of these varieties: if affine varieties X_1 and X_2 over an algebraically closed field k have coordinate rings R_1 and R_2, respectively, and X is the variety with coordinate ring $R_1 \otimes_k R_2$, then we get bijections

$$X(k) \to \mathrm{Hom}_{k\text{-}Alg}(R_1 \otimes_k R_2, k) \overset{\text{Prop.8.2}}{\to} \mathrm{Hom}_{k\text{-}Alg}(R_1, k) \times \mathrm{Hom}_{k\text{-}Alg}(R_2, k) \to X_1(k) \times X_2(k).$$

We similarly get bijections $X(S) \to X_1(S) \times X_2(S)$, functorially in commutative k-algebras S.

8.2. **Some examples of tensor products of algebras.**

Example 8.5. In this example, all rings are commutative.

(i) For any R-algebra S, we claim that we have an isomorphism $S \otimes_R R[x] \cong S[x]$ of R-algebras (it is implicitly understood that $S[x]$ is an R-algebra via $R \to S \hookrightarrow S[x]$). For this, since $R[x]$ is a free R-module with basis the x^i, the commutativity of the tensor product with direct sum gives us an R-module isomorphism $\varphi : S \otimes_R R[x] \to S[x]$. Since this transports $(s \otimes x^i)(s' \otimes x^j) = (ss' \otimes x^{i+j})$ to $ss'x^{i+j} = sx^i \cdot s'x^j$, this is muliplicative at the level of pure tensors, and hence by distributivity multiplicative in general. Hence it is a ring homomorphism (it takes 1 to 1 as well), and clearly is also an R-algebra homomorphism. This example did not even need S to be commutative.

Another way to see this when S is commutative, is to note that in a homomorphism from $R[x]$ or $S[x]$, the "x can go anywhere", so we get identifications

$$\mathrm{Hom}_{R\text{-}alg}(S[x], S') = \mathrm{Hom}_{R\text{-}alg}(S, S') \times S' \cong \mathrm{Hom}_{R\text{-}alg}(S, S') \times \mathrm{Hom}_{R\text{-}alg}(R[x], S'),$$

realizing $S[x]$ as a coproduct of S and $R[x]$ in the category of commutative R-algebras. When S is commutative, $S \otimes_R R[x] \to S[x]$ is clearly in fact an isomorphism of S-algebras.

(ii) More generally, we similarly have $S \otimes_R R[x_j \mid j \in J] \cong S[x_j \mid j \in J]$ as both R-algebras and as S-modules (or as S-algebras if S is commutative), and hence $S[x_i \mid i \in I] \otimes_R R[x_j \mid j \in J] \cong S[x_i \mid i \in I \sqcup J]$.

(iii) As a special case, $R[x] \otimes_R R[x] \cong R[x, y]$, where the '$x$' on the right-hand side corresponds to the '$x \otimes 1$' on the left, [16] and the 'y' on the right corresponds to the '$1 \otimes x$' on the left. For those of you who have seen some basic algebraic geometry, this corresponds to the fact that the product of the affine line $\mathbb{A}^1$ over k with itself is the affine plane $\mathbb{A}^2$ over k.

(iv) Let S_1, S_2 be R-algebras, and let $I_1 \subset S_1$ be an ideal. Then by the right-exactness of tensor products for modules,

$$(S_1/I_1) \otimes_R S_2 \cong (S_1 \otimes S_2)/\text{image}(I_1 \otimes_R S_2 \to S_1 \otimes_R S_2),$$

as R-modules. Check that it is also an isomorphism of R-algebras. It is immediate that the image of $I_1 \otimes_R S_2 \to S_1 \otimes_R S_2$ is simply the ideal $I_1(S_1 \otimes_R S_2)$, where I_1 is understood to act by mapping to $S_1 \otimes_R S_2$ via $I_1 \hookrightarrow S_1 \to S_1 \otimes_R S_2$.

(v) As an even more special case, let us compute $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$. We have, using (iv),

$$\mathbb{C}\otimes_{\mathbb{R}}\mathbb{C} \cong (\mathbb{R}[x]/(x^2+1))\otimes_{\mathbb{R}}\mathbb{C} \overset{\text{(iv)}}{\cong} \mathbb{C}[x]/(x^2+1) \cong \mathbb{C}[x]/((x-i)(x+i)) \cong \mathbb{C}[x]/(x-i)\times\mathbb{C}[x]/(x+i) \cong \mathbb{C}\times\mathbb{C},$$

where the second step from the last used the Sunzi's theorem, i.e., the Chinese remainder theorem.

Check that the above isomorphism maps $a \otimes b$ to $(ab, \bar{a}b)$: an informal explanation is that the i that occurs in the expansion of a using its real and imaginary parts is the "x" of the $\mathbb{C}[x]/(x^2+1)$, and this x was sent to i in the $\mathbb{C}[x]/(x-i) \cong \mathbb{C}$ factor, and to $-i$ in the $\mathbb{C}[x]/(x+i)$ factor.

Example 8.6. This example generalizes Example 8.5(v). Please go through this carefully, it will be considered an important example for this course.

Let E/F be a finitely generated separable extension of fields. The primitive element theorem says that $E = F[\alpha]$ for some $\alpha \in E$. Let f be the minimal polynomial of α. If K is another field containing F, we have

$$E \otimes_F K \cong (F[x]/(f) \otimes_F K) \overset{\text{Example 8.5(iv)}}{\cong} K[x]/(f) \cong \prod_{i=1}^{r} K[x]/(f_i), \tag{33}$$

where $f = f_1 \dots f_r$ is the factorization of f in $K[x]$ into irreducible polynomials, and we have used the Chinese remainder theorem, as justified by the fact that these factors are pairwise coprime (since E/K is separable). Since each f_i is irreducible, each $K[x]/(f_i)$ is a field extension of K, so we might like to express the $K[x]/(f_i)$ in terms of fields that are some how 'composed of E and K'.

[16] Obviously I am talking informally when I say 'corresponds' but what I mean is clear: $x \otimes 1 \mapsto x$, and $1 \otimes x \mapsto y$, under this isomorphism.

This can be done as follows. Let K^{sep}/K be a separable closure of K: thus, K^{sep}/K is a separable algebraic extension, and every separable polynomial over K splits into linear factors over K^{sep}. In particular, f has $\deg f = [E : F]$-many distinct roots in K^{sep}. There is a bijection

$$\mathrm{Hom}_{F\text{-}Alg}(E, K^{sep}) \overset{\text{bij}}{\to} \{\beta \mid \beta \text{ is a root of } f \text{ in } K^{sep}\}, \tag{34}$$

sending $\sigma : E \hookrightarrow K^{sep}$ [17] to $\beta := \sigma(\alpha)$, which is a root of f since α is. The inverse sends β to the unique embedding $\sigma : E \cong F[x]/(f) \hookrightarrow K^{sep}$ that sends α to β.

But recall from (33) that we are interested in describing the $K[x]/(f_i)$, which are obtained by adjoining roots of f_i. This means that we need to group the roots of f (on the right-hand side of (34)) into those that are roots for a common f_i. For this, note that:

- If $\beta \in K^{sep}$ is a root of some f_i, then the roots of f_i are precisely the $\sigma(\beta)$, as σ ranges over $\mathrm{Gal}(K^{sep}/K)$.
- $\mathrm{Gal}(K^{sep}/K)$ acts on the left-hand side of (34), i.e., on $\mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})$, by composition. This makes (34) equivariant for $\mathrm{Gal}(K^{sep}/K)$: if $\sigma : E \hookrightarrow K^{sep}$ maps to β, so $\sigma(\alpha) = \beta$, then for all $\tau \in \mathrm{Gal}(K^{sep}/K)$, $\tau \circ \sigma(\alpha) = \tau(\beta)$, so $\tau \circ \sigma$ maps to $\tau(\beta)$.

Thus, we get a bijection:

$$\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep}) \to \{f_i \mid 1 \leqslant i \leqslant r\}, \tag{35}$$

sending the $\mathrm{Gal}(K^{sep}/K)$-orbit of $\sigma \in \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})$ to the unique f_i such that $\beta := \sigma(\alpha)$ is a root of f_i. For such σ and $\beta = \sigma(\alpha)$, we have:

$$K[x]/(f_i) \cong K[\beta] = K[\sigma(\alpha)] = \text{ the subfield of } K^{sep} \text{ generated by } K \text{ and } \sigma(E),$$

since $E = F[\alpha]$. In other words, we have described $K[x]/(f_i)$ in terms of $\sigma : E \hookrightarrow K^{sep}$, a representative of the $\mathrm{Gal}(K^{sep}/K)$-orbit in $\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})$ corresponding to f_i under the bijection (35).

Conclusion. By (33), we get an isomorphism of F-algebras (even of K-algebras)

$$E \otimes_F K \to \prod_{\sigma \in [\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})]} K_\sigma, \tag{36}$$

where $[\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})]$ is a set of representatives in $\mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})$ for $\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})$, and for σ in this set $K_\sigma \subset K^{sep}$ denotes the subfield generated by K and $\sigma(E)$ inside K^{sep}, which contains both K and $\sigma(E)$. Explicitly, this isomorphism satisfies:

$$a \otimes b \mapsto (\sigma(a)b)_{\sigma \in [\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})]}.$$

Notice how this generalizes Example 8.5(v). If we change the choice of the representatives $\mathrm{Gal}(K^{sep}/K) \backslash \mathrm{Hom}_{F\text{-}Alg}(E, K^{sep})$, we get isomorphic but possibly different right-hand sides, but in any case different isomorphisms.

[17] Any ring homomorphism between fields is injective.

Exercise 8.7. Show – first using direct computation, and then separately using the description around (36) – that

$$\mathbb{Q}[\sqrt[3]{2}] \otimes_{\mathbb{Q}} \mathbb{Q}[\sqrt[3]{2}] \cong \mathbb{Q}[\sqrt[3]{2}] \times \mathbb{Q}[\sqrt[3]{2}, \omega],$$

where ω is a cube root of unity.

When I taught the second semester algebra earlier, one of the questions in the first midterm was to describe $\mathbb{Q}[\sqrt[3]{2}] \otimes_{\mathbb{Q}} \mathbb{Q}[\sqrt[3]{2}]$ without proof.

Example 8.8. Separability was really crucial to the description in Example 8.6. For instance, suppose that F has characteristic p, and that E/F is a purely inseparable extension of degree p obtained by adjoining a p-th root α of some $a \in F$. Then

$$E \otimes_F E \cong F[x]/(x^p - a) \otimes_F E \cong E[x]/(x^p - a) = E[x]/(x - \alpha)^p \cong E[x]/(x^p),$$

the last step using the "change of variables" mapping x to $x + \alpha$. Thus, unlike in (36), this time $E \otimes_F E \cong E[x]/(x^p)$ has nilpotents, and hence cannot be a product of field extensions of F.

Hopefully, we will discuss field and Galois theory in a later lecture, and see that, among finite field extensions E/F. the separable ones can be characterized as those for which $E \otimes_F K$ is a product of fields, while purely inseparable ones acquire "nilpotents" on tensoring with suitable K (e.g., with an algebraic closure of F).

Exercise 8.9. Let $\mathbb{H}$ be the Hamilton quaternions: as a vector space it is a noncommutative $\mathbb{R}$-algebra with an $\mathbb{R}$-basis written as $\{1, i, j, k\}$, and its multiplication is defined by requiring that 1 is the multiplicative identity and $i^2 = j^2 = k^2 = -1, ij = k, jk = i$ and $ki = j$. Then show that $\mathbb{H} \otimes_R \mathbb{C}$ (which is clearly a $\mathbb{C}$-algebra) is isomorphic to $\mathrm{M}_2(\mathbb{C})$ as a $\mathbb{C}$-algebra.

8.3. **Tensor algebras.** Henceforth, till the end of this lecture, R will denote an arbitrary commutative ring, unless otherwise specified. Let ${}^{nc}R\text{-}Alg$ denote the category of noncommutative (i.e., not necessarily commutative) R-algebras, and $R\text{-}Alg$ the category of commutative R-algebras.

A motivating question is: does

$$Forget : {}^{nc}R\text{-}Alg \rightsquigarrow R\text{-}Mod$$

have a left adjoint? Namely, does it make sense to talk of a "free R-algebra on a given R-module M?"

To motivate, let us look for an R-algebra T together with functorial bijections

$$\mathrm{Hom}_{{}^{nc}R\text{-}Alg}(T, S) \overset{\text{bij}}{\to} \mathrm{Hom}_R(M, S)$$

for all noncommutative R-algebras S and R-modules M. Let S be an R-algebra, say $\varphi_0 : R \to S$, together with an R-module homomorphism $\varphi_1 : M \to S$.

We then have, for all $r \geqslant 1$, an r-multilinear map

$$M \times \cdots \times M \to S, \quad (m_1, \ldots, m_r) \mapsto \varphi_1(m_1) \ldots \varphi_1(m_r),$$

which therefore quotients to a map $\varphi_r : M^{\otimes^r} \to S$, where $T^r(M) = M^{\otimes r}$ stands for the n-fold tensor product $M \otimes \cdots \otimes M$ of M with itself. We set $M^{\otimes 0} = T^0(M) = R$.

Therefore, we get a map

$$T(M) := \bigoplus_{r \geqslant 0} \varphi_r : \bigoplus_{r \geqslant 0} M^{\otimes r} \to S. \tag{37}$$

It is immediate how to turn the left-hand side into a ring: for $r, s \geqslant 1$, we have an isomorphism

$$M^{\otimes\, r} \otimes_R M^{\otimes\, s} \to M^{\otimes (r+s)},$$

by an easy argument generalizing the isomorphism $(M \otimes_R N) \otimes_R L \cong M \otimes_R N \otimes_R L$ from Lecture 6. This defines a bilinear map $M^{\otimes\, r} \times M^{\otimes\, s} \to M^{\otimes (r+s)}$, and extending by distributivity we get a map

$$T(M) \times T(M) \to T(M).$$

Definition 8.10. (i) $T(M)$, which will be seen in Exercise 8.11 below to be an R-algebra, is called the tensor algebra associated to the R-algebra M.

(ii) By the functoriality of the tensor product discussed in Lecture 6, $T^r = \otimes^r$, for each $r > 0$, is a functor $R\text{-}Mod \rightsquigarrow R\text{-}Mod$, where for $f \in \mathrm{Hom}_R(M, N)$, $T^r(f) = \otimes^r f : \otimes^r M \to \otimes^r N$. For $r = 0$, we view T^0 as a functor by defining, for each homomorphism $f : M \to N$ of R-modules, $T^0(f) : T^0(M) = R \to R = T^0(N)$ to be the identity. Thus, now we have a functor $T^r(-) : R\text{-}Mod \rightsquigarrow R\text{-}Mod$, for each $r \geqslant 0$.

(iii) $T(-) : R\text{-}Mod \rightsquigarrow {}^{nc}R\text{-}Alg$ will stand for the functor that assigns to an R-module M its tensor algebra $T(M)$, and to an R-module homomorphism $f : M \to N$ the map $T(f) = \bigoplus_{r=0}^{\infty} T^r(f) : T(M) \to T(N)$, which is an algebra homomorphism by Exercise 8.11(ii) below.

Exercise 8.11. (i) Show that the above map $T(M) \times T(M) \to T(M)$ satisfies associativity, and that for the multiplication it defines, $T(M)$ has an identity given by with $1 \in R = T^0 M \hookrightarrow T(M)$: thus, $T(M)$ is a ring. In fact, since $R = \otimes^0 M \hookrightarrow T(M)$ has image in the center of $T(M)$, it follows that $R \hookrightarrow T(M)$ is in fact an R-algebra.

(ii) Prove the claim in Definition 8.10(iii): if $M \to N$ is an R-module homomorphism, then $\bigoplus_{r=0}^{\infty} T^r(f) : T(M) \to T(N)$ is a homomorphism of R-algebras.

Proposition 8.12. *For an R-module M, consider the R-module homomorphism $\iota_M : M = T^1(M) \hookrightarrow T(M)$. Then for any $M \in \mathrm{Ob}\, R\text{-}Mod$ and $S \in \mathrm{Ob}\, {}^{nc}R\text{-}Alg$, precomposition with ι_M induces a bijection:*

$$\mathrm{Hom}_{{}^{nc}R\text{-}Alg}(T(M), S) \overset{-\circ \iota_M}{\to} \mathrm{Hom}_R(M, S),$$

realizing $T(-)$ as left-adjoint to the forgetful functor ${}^{nc}R\text{-}Alg \rightsquigarrow R\text{-}Mod$.

Proof. An exercise, using the above discussion. □

Exercise 8.13. (i) Define what a 'noncommutative polynomial algebra in n variables' should mean. If $M = R^n$, show that $T(M)$ is a noncommutative polynomial algebra in n variables over R.

(ii) When $M = R/I$ for an ideal I, show that we have an identification

$$T(M) \cong \{a_0 + a_1 x + \cdots + a_n x^n \mid n \in \mathbb{N}, a_0 \in R, a_i \in R/I \,\forall\, 1 \leqslant i \leqslant n\},$$

with addition and multiplication defined just as for polynomials.

$T(M)$ is not just an R-algebra, but an R-algebra graded by the additive monoid $\mathbb{N} := \mathbb{Z}_{\geqslant 0}$ (recall that for us, 0 is a natural number):

Definition 8.14. (i) A ring T is said to be graded by a monoid G, or G-graded, if we have a decomposition of abelian groups:

$$T = \bigoplus_{g \in G} T_g, \tag{38}$$

such that $T_g \cdot T_h \subset T_{gh}$ for all $g, h \in G$. The various T_g are called the graded or homogeneous components of T; T_g is said to be the graded or homogeneous component of degree g.

(ii) We say that T is a G-graded R-algebra if T is both a G-graded ring and an R-algebra, such that the image of $R \to T$ lies in the graded component T_e corresponding to the identity element $e \in G$. Thus, if T is a G-graded R-algebra, then we have a decomposition as in (38), but involving R-modules in place of abelian groups. [18]

(iii) By homogeneous elements of T, we refer to elements that belong to T_g for some $g \in G$. If $x \in T_g$, we say that x is homogeneous of degree d.

(iv) By just graded, without specifying a G, we will mean $\mathbb{N}$-graded.

Remark 8.15. In the lecture, I gave a different and incorrect definition of a graded R-algebra, as opposed to the standard one given above. One of you called out my definition: I had simply required (38) to be a decomposition of R-modules. This requirement is satisfied if T is a graded R-algebra in the 'standard' sense given above: if $R \to T$ lands in T_e, then (38) is indeed a decomposition of R-modules. However, the converse may not be true. Nevertheless, the converse does seem clear if the monoid G satisfies that $h \neq gh \neq g$ for all $e \neq g, h \in G$: indeed, in this case the projection of $1 \in T$ to T_e under (38) also functions as a multiplicative identity, forcing $1 \in T_e$, and therefore also that $R \subset T_e$. So for $\mathbb{N}$-graded rings, the definition I gave should also work. But the standard definition is a better one, as it certainly seems to work better for general monoids.

To see that $T(M)$ is an $\mathbb{N}$-graded R-algebra, use that $T^r(M) \cdot T^s(M) \subset T^{r+s}(M)$, and note that $R = T^0(M)$. This lets us view $T(-)$ as a functor

$$R\text{-}Mod \rightsquigarrow {}^{nc}Gr\text{-}R\text{-}Alg,$$

where ${}^{nc}Gr\text{-}R\text{-}Alg$ stands for the category of $\mathbb{N}$-graded not necessarily commutative R-algebras.

[18] where, as usual, the R-module structure on T comes from the R-algebra structure on T.

Exercise 8.16. Show that T, viewed as a functor $R\text{-}Mod \rightsquigarrow {}^{nc}Gr\text{-}R\text{-}Alg$, is left adjoint to $Forget : {}^{nc}Gr\text{-}R\text{-}Alg \rightsquigarrow R\text{-}Mod$ (the proof of Proposition 8.12 is respectful of grading, and hence goes through for this case).

The following simple exercises are kind of boring, but they will be needed for their analogues for the symmetric and exterior algebras.

Exercise 8.17. (i) If $B : M \times N \to R$ is an R-bilinear pairing of R-modules M and N, show that there exists a unique pairing

$$B^{\otimes n} : M^{\otimes n} \times N^{\otimes n} \to R,$$

such that for all $x_1, \ldots, x_n \in M$ and $y_1, \ldots, y_n \in N$, we have

$$B^{\otimes n}(x_1 \otimes \cdots \otimes x_n, y_1 \otimes \cdots \otimes y_n) = \prod_{i=1}^{n} B(x_i, y_i).$$

(ii) If $M = M_1 \oplus M_2$, show, using the maps $T(M_1) \to T(M)$ and $T(M_2) \to T(M)$, an isomorphism, for all $n \in \mathbb{N}$:

$$T^n(M) \cong \bigoplus_{f:\{1,\ldots,n\}\to\{1,2\}} M_{f(1)} \otimes \cdots \otimes M_{f(n)} \cong \bigoplus_{\substack{p,q\in\mathbb{N}\\ p+q=n}} (M_1^{\otimes p} \otimes M_2^{\otimes q})^{\binom{n}{p}}.$$

8.4. **The definition of symmetric and exterior algebras.** Symmetric and exterior algebras are quotients of tensor algebras: the former are 'the commutative variant' of tensor algebras, and the latter are a 'graded commutative' or 'super' version of the tensor algebra, a notion which is important but which we will not discuss much.

Recall that for a noncommutative ring T and a two-sided ideal $I \subset T$, the quotient T/I is a ring (this wouldn't be true if I were only a left-ideal or a right-ideal). If T is a graded ring, when does T/I get a grading from T?

Definition 8.18. (i) Let $T = \bigoplus_{g\in G} T_g$ be a noncommutative ring graded by a monoid G. A left, right or two-sided ideal $I \subset T$ is said to be a homogeneous ideal if it is generated by homogeneous elements. Equivalently (prove this equivalence as an easy exercise), if the inclusion

$$\bigoplus_{g\in G}(I \cap T_g) \subset I$$

is an equality. A similar result applies to graded R-algebras.

(ii) If $T = \bigoplus_{g\in G} T_g$ is a noncommutative ring graded by a monoid G, and $I = \bigoplus_{g\in G} I_g$ is a homogeneous two-sided ideal of T, where $I_g = I \cap T_g$, then T/I has an obvious grading:

$$T/I = \bigoplus_{g\in G} T_g/I_g$$

(note that $(T_g/I_g) \cdot (T_h/I_h) \subset T_{gh}/I_{gh}$), and hence will be viewed as a graded ring.

Now we can define symmetric and exterior algebras:

Definition 8.19. (i) (a) For any R-module M, the symmetric algebra $S(M)$ of M is the quotient of $T(M)$ by the two-sided ideal $I_S(M) \subset T(M)$ generated by $\{x \otimes y - y \otimes x \mid x, y \in M\} = T^1(M) \subset T(M)$. Since each $x \otimes y - y \otimes x$ is homogeneous of degree 2, $I_S(M)$ is a homogeneous ideal, so that $S(M)$ is an $\mathbb{N}$-graded R-algebra:

$$S(M) = \bigoplus_{n \geqslant 0} S^n(M).$$

The n-th graded piece $S^n(M)$ of $S(M)$ with its obvious structure of an R-module, will be called the n-th symmetric power of M.

(b) For any R-module M, the exterior algebra $\Lambda(M)$ of M is the quotient of $T(M)$ by the two-sided ideal $I_\Lambda(M) \subset T(M)$ generated by $\{x \otimes x \mid x \in T^1(M) \subset T(M)\}$. Since each $x \otimes x$ is homogeneous of degree 2, $\Lambda(M)$ is an $\mathbb{N}$-graded R-algebra:

$$\Lambda(M) = \bigoplus_{n \geqslant 0} \Lambda^n(M).$$

The n-th graded piece $\Lambda^n(M)$ of $\Lambda(M)$, with its obvious structure of an R-module, will be called the n-th exterior power of M.

(ii) Given any homomorphism $f : M \to N$ of R-modules, note that $T(f) : T(M) \to T(N)$ sends $I_S(M)$ to $I_S(N)$ and $I_\Lambda(M)$ to $I_\Lambda(N)$, and hence descends to R-algebra homomorpisms $S(f) : S(M) \to S(N)$ and $\Lambda(f) : \Lambda(M) \to \Lambda(N)$. This defines functors $S(-), \Lambda(-) : R\text{-}Mod \rightsquigarrow {}^{nc}Gr\text{-}R\text{-}Alg$.

(iii) Since $I_S(M)$ and $I_\Lambda(M)$ consist of elements of degree at least two, the maps

$$\begin{cases} R \to T^0(M) \to S^0(M) \\ R \to T^0(M) \to \Lambda^0(M) \end{cases}, \quad \text{and} \quad \begin{cases} M \to T^1(M) \to S^1(M) \\ M \to T^1(M) \to \Lambda^1(M) \end{cases}$$

are isomorphisms, giving in particular injections $\iota_{S,M} : M \hookrightarrow S(M)$ and $\iota_{\Lambda,M} : M \hookrightarrow \Lambda(M)$.

(iv) Given $m_1 \otimes \cdots \otimes m_n \in T^n(M) \subset T(M)$, we denote its image in $S^n(M)$ by $m_1 \dots m_n$, and its image in $\Lambda^n(M)$ by $m_1 \wedge \cdots \wedge m_n$.

8.5. **Some basic properties of symmetric powers.** We will usually write $M^{\times n}$ for the n-fold product $\times_{i=1}^n M = M \times \cdots \times M$ (we usually write this instead of M^n when we wish to discuss multilinear maps from it; although this is probably non-standard notation). Recall that $R\text{-}Alg$ is the category of commutative R-algebras.

For each $n \geqslant 0$, note that the symmetric group $\mathfrak{S}_n$ on n letters [19] acts by permutation on $\times_{i=1}^n M$, pulling multilinear forms back to multilinear forms. Hence this gives an action of $\mathfrak{S}_n$ on $T^n(M) = M^{\otimes n}$ by R-module automorphisms, which satisfies: $\sigma \cdot (m_1 \otimes \cdots \otimes m_n) = (m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(n)})$.

The following long exercise studies the symmetric algebra and the symmetric powers:

[19] Using $\mathfrak{S}_n$ instead of S_n for the symmetric group, to avoid confusion with the S of $S(M)$.

Exercise 8.20. (i) $I_S(M) \cap T^n(M) \subset T^n(M) = M^{\otimes n}$ is the span of all the

$$(m_1 \otimes \cdots \otimes m_i \otimes \cdots \otimes m_j \otimes \cdots \otimes m_n) - (m_1 \otimes \cdots \otimes m_j \otimes \cdots \otimes m_i \otimes \cdots \otimes m_n),$$

as the $(m_1, \ldots, m_n) \in M^{\times n}$ and the $1 \leqslant i, j \leqslant n$ vary.

To put it another way, show that $I_S(M) \subset T^n(M)$ is the span of all $\sigma \cdot x - x$, where σ varies over $\mathfrak{S}_n$ and x over $T^n(M)$. [20]

(ii) Deduce as a consequence that the map $u : M^{\times n} \to S^n(M)$ has the following universal property: u is multilinear and symmetric, in the sense that it takes the same value on $(m_1, \ldots, m_n)$ and $(m_{\sigma(1)}, \ldots, m_{\sigma(n)})$ for each $(m_1, \ldots, m_n) \in M^{\times n}$ and each $\sigma \in \mathfrak{S}_n$, and for any R-module L, $- \circ u$ induces a bijection

$$\mathrm{Hom}_R(S^n(M), L) \xrightarrow{-\circ u} \Big\{\text{Symmetric multilinear maps } M^{\times n} \to L\Big\}.$$

(iii) Show that $S(M)$ is a commutative R-algebra.
Hint: The image of $\iota_{S,M} : M \hookrightarrow S(M)$ generates $S(M)$, and the definition imposes commutativity on these generators.

(iv) Show that for all commutative R-algebras A and R-modules M, composition with $\iota_{S,M} : M \hookrightarrow S(M)$ gives a bijection,

$$\mathrm{Hom}_{R\text{-}Alg}(S(M), A) \xrightarrow{-\circ \iota_{S,M}} \mathrm{Hom}_R(M, A).$$

This realizes $S(-)$ as a left-adjoint to the forgetful functor $R\text{-}Alg \rightsquigarrow R\text{-}Mod$.

(v) $S(-)$ can also be viewed as a functor from $R\text{-}Mod$ into the category $Gr\text{-}R\text{-}Alg$ of graded commutative R-algebras. Show that $- \circ \iota_{S,M}$ also realizes $S(-)$ as a left-adjoint to the forgetful functor $Gr\text{-}R\text{-}Alg \rightsquigarrow R\text{-}Mod$.

(vi) Let M be a free R-module with basis $\{e_s \mid s \in S\}$. Then, viewing $\iota_{S,M} : M \hookrightarrow S(M)$ as an inclusion, show that we have an isomorphism $S(M) \cong R[x_s \mid s \in S]$, sending each e_s to the variable x_s. In particular, $S^n(M)$ consists of the various monomials $e_{s_1} \ldots e_{s_n}$ of degree n, with each $s_i \in S$.

Thus, if M is a R-free module of rank r, then $S^n(M)$ is a free R-module of rank $\binom{r+n-1}{n}$.

(vii) Read and convince yourself of the following. The above important exercise tells you that the symmetric algebra on a free module is a coordinate-free version of a polynomial algebra. It also tells you how to think of polynomials in a coordinate-free fashion: if V is a free R-module, and $V^\vee := \mathrm{Hom}_R(V, R)$, then:

$$S(V^\vee) \text{ is the space of polynomial functions on } V.$$

More precisely elements of $S^0(V^\vee) = R$ are the constant functions on V, the elements of $V^\vee = \mathrm{Hom}_R(V, R)$ are the linear functions on V without a constant term, the elements of $S^2(V^\vee)$ are the homogeneous quadratic functions on V – they are the linear combinations of products of elements of $V^\vee$ – and so on.

There is a caveat here: if R is 'small', different elements of $S(V^\vee)$ may give the same function on V; but if you interpret the meaning of a 'polynomial function'

[20] This way of quotienting a module N on which $\mathfrak{S}_n$ acts, by the span of all the $\sigma(x) - x$ as x varies over N, is called the process of taking coinvariants.

suitably – giving a map $V \otimes_R S \to S$ for each commutative R-algebra S, and functorially so – the above can still be viewed as making sense. In fact, this is how polynomial maps are viewed in a lot of algebraic geometry.

(viii) This exercise describes an isomorphism

$$S(M_1) \otimes_R S(M_2) \cong S(M_1 \oplus M_2)$$

of R-algebras. Suppose $M = M_1 \oplus M_2$ is the direct sum of two R-modules. Applying $S(-)$ to $M_1 \hookrightarrow M$ and $M_2 \hookrightarrow M$, we get homomorphisms $S(M_1) \to S(M)$ and $S(M_2) \to S(M)$ of graded R-algebras, and hence – using that the tensor product is a coproduct in the category of commutative R-algebras (which these rings are, by (iii) above) – a homomorphism of R-algebras,

$$S(M_1) \otimes_R S(M_2) \to S(M).$$

Show that this homomorphism is an isomorphism.

Hint: This is just because left-adjoints commute with coproducts.

(ix) Conclude from the previous exercise that, when $M = M_1 \oplus M_2$ as in that exercise, we have for each $n \in \mathbb{N}$, an isomorphism:

$$S^n(M) \cong \bigoplus_{\substack{p,q \in \mathbb{N} \\ p+q=n}} S^p(M_1) \otimes_R S^q(M_2),$$

whose inverse is obtained using $S^p(M_1 \to M)$ and $S^q(M_2 \to M)$ together with the map $S^p(M) \otimes_R S^q(M) \to S^{p+q}(M)$ given by multiplication in $S(M)$.

Hint: Define the notation of a 'graded tensor product', so that $S(M') \otimes_R S(M'')$ naturally has a grading, and show that $S(M') \otimes_R S(M'') \to S(M)$ respects the grading.

(x) *Symmetric tensors.* By symmetric tensors in $M^{\otimes n}$, we mean $(M^{\otimes n})^{\mathfrak{S}_n}$: While $S^n(M)$ is only a *quotient module* of $T^n(M)$, the symmetric tensors form a *submodule* of $T^n(M)$.

We can define the symmetrization map from $M^{\otimes n}$ to $(M^{\otimes n})^{\mathfrak{S}_n}$, by :

$$x \mapsto \sum_{\sigma \in \mathfrak{S}_n} \sigma \cdot x$$

namely, summing over the group action. If $n!$ is invertible in R, show that this map, or equivalently the 'averaging map'

$$x \mapsto (n!)^{-1} \sum_{\sigma \in \mathfrak{S}_n} \sigma \cdot x,$$

defines an isomorphism from $(M^{\otimes n})^{\mathfrak{S}_n}$ to $S^n(M)$. Thus, in this case (i.e., when $n!$ is invertible in R), we can use this map to think of $S^n(M)$ as the space of symmetric tensors.

Hint: Prove this in greater generality, which makes it easier to see what is going on: if G is a finite group with a map $G \to \mathrm{Aut}_R(M)$, where M is an R-module,

and if $\#G$ is invertible in R, then show that

$$m \mapsto (\#G)^{-1} \sum_{g \in G} g \cdot m,$$

induces an isomorphism $M_G \to M^G$, where

$$M_G := M/\mathrm{Span}_{\mathbb{Z}}(\{g \cdot m - m \mid g \in G, m \in M\})$$

is the quotient module of G-coinvariants for M, and $M^G \subset G$ is the submodule consisting of G-fixed elements.

(xi) (This exercise doesn't seem that important, but its exterior power analogue will be important).
If $B : M \times N \to R$ is an R-bilinear pairing of R-modules M and N, show that there exists a unique pairing

$$S^n(B) : S^n(M) \times S^n(N) \to R,$$

such that for all $x_1, \ldots, x_n \in M$ and $y_1, \ldots, y_n \in N$, we have

$$S^n(B)(x_1 \ldots x_n, y_1 \ldots y_n) = \sum_{\sigma \in S_n} B^{\otimes n}(x_1 \otimes \cdots \otimes x_n, \sigma(y_1 \otimes \cdots \otimes y_n)) = \sum_{\sigma \in S_n} B(x_1, y_{\sigma(1)}) \ldots B(x_n, y_{\sigma(n)}),$$

with $B^{\otimes n}$ as in Exercise 8.17(i).

If further M and N are free with bases $e_1, \ldots, e_r$ and $f_1, \ldots, f_r$, respectively, with $B(e_i, f_j) = \delta_{i,j}$ (thus, N is dual to M), show that this pairing between $S^n(M)$ and $S^n(N)$ can be given as follows: given basis elements $e_I = e_{i_1} \ldots e_{i_n}$ of $S^n(M)$ and $f_J = f_{j_1} \ldots f_{j_n}$ of $S^n(N)$, where I and J are multisets over $\{1, \ldots, r\}$, [21] then $S^n(B)(e_I, e_J) = 0$ if $I \neq J$, and equals $\prod_{j=1}^r n(j, I)!$ otherwise, where $n(j, I)$ is the number of elements in I that equal J.

Thus, this pairing is somewhat ugly, and even when M are N are free modules of finite rank in duality with each other it is only perfect if $n! \in R^\times$.

8.6. **Some basic properties of exterior powers.**

Remark 8.21. In this subsection, we will need the notion of the sign $\mathrm{sgn}(\sigma) \in \{\pm 1\}$ of a permutation $\sigma \in \mathfrak{S}_n$. The following are equivalent ways to define $\mathrm{sgn}(\sigma)$:

- If $\sigma = s_1 \ldots s_m$ with each $s_n \in \mathfrak{S}_n$ a transposition, then $\mathrm{sgn}(\sigma) = (-1)^m$. The reason this is well-defined is that, if $\sigma = s'_1 \ldots s'_{m'}$ is another such decomposition, then one can show that $m - m'$ is even. One way to prove this is to use the other descriptions in the next point.
- Represent σ as a matrix: say $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$, and σ can be represented by the permutation matrix that sends each e_i to $e_{\sigma(i)}$. The determinant of this matrix is $\mathrm{sgn}(\sigma)$ as discussed in the previous point: this follows from the fact that the permutation matrix associated to each transposition clearly has determinant -1.

[21] "multisets" are like sets but with repetition allowed; so here I and J contain elements from $1, \ldots, r$, but they are not sequences, in that we don't keep track of the order

- $\mathrm{sgn}(\sigma)$ equals $(-1)^N$, where $N = \#\{(x,y) \mid x < y \text{ and } \sigma(x) > \sigma(y)\}$. You can try to prove this or look up somewhere; we will not use this today.

The following exercise studies the exterior algebra and the exterior powers:

Exercise 8.22. (i) Show the following facts about $I_\Lambda(M)$:

(a) $I_\Lambda(M)$ contains $x \otimes y + y \otimes x$ for all $x, y \in M$. Thus, the 'alternating condition' of annihilating every $x \otimes x$ implies the 'skewsymmetry condition' of annihilating every $x \otimes y + y \otimes x$.
Hint: Consider $(x+y) \otimes (x+y)$, $x \otimes x$ and $y \otimes y$.

(b) If 2 is invertible in R, then the two-sided ideal generated by the $x \otimes y + y \otimes x$ contains all the $x \otimes x$, and hence contains $I_\Lambda(M)$. Thus, when 2 is invertible, the skewsymmetry condition implies the alternating condition.

(ii) (a) $I_\Lambda(M) \cap T^n(M) \subset T^n(M) = M^{\otimes n}$ can be described as the span of all the the elements of the form $m_1 \otimes \cdots \otimes m_n$, such that $m_i = m_{i+1}$ for some $1 \leqslant i < n$.

(b) $I_\Lambda(M) \cap T^n(M)$ contains all elements of the form

$$m_1 \otimes \cdots \otimes m_i \otimes \cdots \otimes m_j \otimes \cdots \otimes m_n + m_1 \otimes \cdots \otimes m_j \otimes \cdots \otimes m_i \otimes \cdots \otimes m_n.$$

Hint: First consider the case where $j = i+1$. Then iterate what you get an odd number of times for more general j. If this looks too complicated, first consider $n = 3$.

(c) $I_\Lambda(M) \cap T^n(M) \subset T^n(M) = M^{\otimes n}$ can also be described as the span of all the the elements of the form $m_1 \otimes \cdots \otimes m_n$ with $m_i = m_j$ for some $1 \leqslant i, j \leqslant n$ with $i \neq j$.
Hint: Use the above part.

(d) If 2 is invertible in R, then $I_\Lambda(M) \cap T^n(M) \subset M^{\otimes n}$ can also be described as the span of all the elements of the form

$$m_1 \otimes \cdots \otimes m_i \otimes \cdots \otimes m_j \otimes \cdots \otimes m_n + m_1 \otimes \cdots \otimes m_j \otimes \cdots \otimes m_i \otimes \cdots \otimes m_n.$$

To put it another way, show that if 2 is invertible in R, then $I_\Lambda(M) \subset T^n(M)$ is the span of all $\sigma \cdot x - \mathrm{sgn}(\sigma) \cdot x$, where σ varies over $\mathfrak{S}_n$ and x over $T^n(M)$.[22]

(iii) Deduce from (c) of the previous problem that the map $u : M^{\times n} \to \Lambda^n(M)$ has the following universal property: u is multilinear and alternating, where 'alternating' means that it vanishes on $(m_1, \ldots, m_n)$ whenever $m_i = m_j$ for some $i \neq j$, and for any R-module L, $- \circ u$ induces a bijection

$$\mathrm{Hom}_R(\Lambda^n(M), L) \xrightarrow{-\circ u} \Big\{\text{Alternating multilinear maps } M^{\times n} \to L\Big\}.$$

[22] This is again a process of taking coinvariants, but this time "$(\mathfrak{S}_n, \chi)$-coinvariants": if a group G acts on an R-module N and $\chi : G \to R^\times$ is a character, then the R-module N_χ of (G,χ)-coinvariants of N is the quotient of N by the span of all the $g \cdot n - \chi(g)n$ such that $g \in G$ and $n \in N$. Then N_χ is a quotient R-module of N, and gets an induced action of G: the induced action of each $g \in G$ on N_χ is then simply by multiplication by $\chi(g)$. Moreover, any G-equivariant R-module homomorphism from N to another module on which G acts through χ factors through $N \to N_\chi$. Thus, this generalizes the discussion of coinvariants we had in the context of $S^n(M)$.

(iv) Show that $\Lambda(M)$ is a graded commutative R-algebra, where 'graded commutative' means that for all $x \in \Lambda^r(M)$ and $y \in \Lambda^s(M)$,

$$x \wedge y = (-1)^{rs} y \wedge x.$$

Hint: The image of $\iota_{\Lambda,M} : M \hookrightarrow \Lambda(M)$ generates $\Lambda(M)$, and the definition imposes graded commutativity on these generators.

(v) Show that for all not necessarily commutative R-algebras A and R-modules M, composition with $\iota_{\Lambda,M} : M \hookrightarrow \Lambda(M)$ gives a bijection,

$$\{\varphi \in \mathrm{Hom}_{^{nc}R\text{-}Alg}(\Lambda(M), A) \mid \varphi(x)^2 = 0 \,\forall\, x \in M\} \overset{-\circ\iota_{\Lambda,M}}{\to} \mathrm{Hom}_R(M, A).$$

One can interpret this as implying that $\Lambda(-)$ is left-adjoint to the forgetful functor $R\text{-}superAlg \rightsquigarrow R\text{-}Mod$, where $R\text{-}superAlg$ is the category of super R-algebras, but let us not worry about this.

(vi) If M is generated as an R-module by $e_1, \ldots, e_r$, show that $\Lambda^n M$ is generated by the various $e_{i_1} \wedge \cdots \wedge e_{i_n}$, where the $i_1, \ldots, i_n$ ranges over *strictly* increasing sequences of numbers between 1 and r. Deduce that $\Lambda^n M = 0$ if $n > r$.

(vii) This is not trivial, but is somewhat important; please make sure to look at this. Suppose M is a free R-module with basis $e_1, \ldots, e_r$. Show that $\Lambda^n M$ is a free R-module with basis consisting of the various $e_{i_1} \wedge \cdots \wedge e_{i_n}$, where the sequence $i_1, \ldots, i_n$ runs over *strictly* increasing sequences of numbers between 1 and n. Thus, it is a free R-module of rank $\binom{r}{n}$.
Hint: By the previous problem, it suffices to show that the $e_{i_1} \wedge \cdots \wedge e_{i_n}$ are linearly independent. Define $M^{\otimes n} \to M^{\otimes n}$ by an 'antisymmetrization map':

$$x \mapsto \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma)(\sigma \cdot x).$$

This map vanishes on $I_\Lambda(M) \cap T^n(M)$, and hence factors through $\Lambda(M)$. This factored map sends $e_{i_1} \wedge \cdots \wedge e_{i_n}$ to $\sum_\sigma \mathrm{sgn}(\sigma) e_{\sigma(i_1)} \otimes \ldots e_{\sigma(i_n)}$. Now for two distinct such sequences $i_1 < \cdots < i_n$ and $j_1 < \cdots < j_n$, and $\sigma, \tau \in \mathfrak{S}_n$, $i_{\sigma(1)}, \ldots, i_{\sigma(n)}$ and $j_{\tau(1)}, \ldots, j_{\tau(n)}$ are distinct, since we started with distinct increasing sequences. Thus, the various $\sum_\sigma \mathrm{sgn}(\sigma) e_{\sigma(i_1)} \otimes \ldots e_{\sigma(i_n)}$ as above involve coefficients from disjoint sets of basis elements of $M^{\otimes n}$. This forces them to be linearly independent.

(viii) Suppose $M = M_1 \oplus M_2$ is the direct sum of two R-modules. Applying $\Lambda(-)$ to $M_1 \hookrightarrow M$ and $M_2 \hookrightarrow M$, we get homomorphisms $\Lambda(M_1) \hookrightarrow \Lambda(M)$ and $\Lambda(M_2) \hookrightarrow \Lambda(M)$ of R-algebras. Show that this induces an isomorphism

$$\Lambda^n(M) \cong \bigoplus_{\substack{p,q \in \mathbb{N} \\ p+q=n}} \Lambda^p(M_1) \otimes_R \Lambda^q(M_2).$$

Like with symmetric algebras, we can interpret this as an equality as saying that $\Lambda(M)$ is an appropriate graded tensor product of $\Lambda(M_1)$ and $\Lambda(M_2)$. There is a way to adapt the proof from the case of symmetric algebras. But let us not worry about this; one can do this directly.

Hint: The inverse of the isomorphism asked for has been constructed in the problem. Construct an explicit candidate for the isomorphism itself: given an element of $\Lambda^n(M)$ which is a wedge of a particular sequence of elements from M_1 and M_2, use a permutation σ to have all the M_1-terms first, and then all the M_2-terms (somewhat in the spirit of Exercise 8.17(ii)), but compensate for this by throwing in a $\mathrm{sgn}(\sigma)$.

(ix) *Application to the structure theorem for modules over a PID.* Use the above exercise to show that if

$$M = R/(d_1) \oplus \cdots \oplus R/(d_r),$$

with $d_r | \ldots | d_1$ (for convenience we have inverted the usual order), then for $j \geqslant 1$ we have:

$$\Lambda^j(M) = \begin{cases} \bigoplus_{1 \leqslant i_1 < \cdots < i_j \leqslant r} R/(d_{i_j}), & \text{if } j \leqslant r, \text{ and} \\ 0, & \text{if } j > r. \end{cases}$$

the annihilator $\mathrm{Ann}_R(\Lambda^j M)$ of $\Lambda^j M$ in R is d_j. Deduce from this another proof of the uniqueness assertion in the structure theorem for modules over a PID.

Hint: The above problem gives

$$\Lambda^j(M_1 \oplus \cdots \oplus M_r) = \bigoplus_{\substack{i_1,\ldots,i_r \geqslant 0 \\ i_1 + \cdots + i_r = j}} \Lambda^{i_1}(M_1) \otimes \cdots \otimes \Lambda^{i_r}(M_r).$$

Further, by (vi) above, $\Lambda^i(R/I) = 0$ if $i \geqslant 2$ and $I \subset R$ is any ideal. This gives

$$\Lambda^j(M) = \bigoplus_{\{s_1,\ldots,s_j\} \subset \{1,\ldots,r} R/(d_{s_1}) \otimes_R \cdots \otimes_R R/(d_{s_j}).$$

Finally, if $s_1 < \cdots < s_j$, then $d_{s_j} | \ldots | d_{s_1}$, and it is easy to see from the right-exactness of the tensor product that $R/(d_{s_1}) \otimes_R \cdots \otimes_R R/(d_{s_j}) \cong R/(d_{s_j})$, which is a quotient of $R/(d_j)$. Put all these together (this is not all of the solution, but quite close to it. If you are not able to do this, please make sure to ask me).

(x) This will be useful later when we discuss quadratic forms, probably in Lecture 9.

(a) If $B : M \times N \to R$ is an R-bilinear pairing of R-modules M and N, show that there exists a unique pairing

$$\Lambda^n B : \Lambda^n M \times \Lambda^n N \to R,$$

such that for all $x_1, \ldots, x_n \in M$ and $y_1, \ldots, y_n \in N$, we have

$$\Lambda^n B(x_1 \wedge \cdots \wedge x_n, y_1 \wedge \cdots \wedge y_n) = \det([B(x_i, y_j)]_{1 \leqslant i,j \leqslant n}) = \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{j=1}^{n} B(x_i, y_{\sigma(j)}).$$

by means of the abovge pairing.

(b) Let $e_1, \ldots, e_r$ be a basis for a free R-module M, and let $e_1^\vee, \ldots, e_r^\vee$ be the dual basis for $M^\vee$. Apply (a) above to the natural pairing $M^\vee \times M \to R$, and conclude that the resulting pairing $\Lambda^n M^\vee \times \Lambda^n M \to R$ realizes $\Lambda^n(M)^\vee$ as isomorphic to $\Lambda^n(M^\vee)$. More precisely, recalling from (vii) above that $\Lambda^n M$ and $\Lambda^n M^\vee$ have bases consisting of the $e_{i_1} \wedge \cdots \wedge e_{i_n}$ and the $e_{i_1}^\vee \wedge \cdots \wedge e_{i_n}^\vee$,

where $i_1, \ldots, i_n$ run over strictly increasing sequences of elements of $1, \ldots, r$, show that these bases are dual to each other for the pairing between $\Lambda^n M^\vee$ and $\Lambda^n M$ described above.

(xi) Let M be a free module of finite rank r. This problem introduces some basic linear algebra applications, such as a coordinate-free treatment of the determinant and minors of a matrix.

(a) By (vii) above, $\Lambda^r M$ is a free rank one module. Given any $T \in \mathrm{End}_R(M)$, therefore, $\Lambda^r T : \Lambda^r M \to \Lambda^r M$ (as made sense of using the functoriality of Λ^r) is an endomorphism of a free rank one module, and is hence given by multiplication by a scalar, called $\det T$. [23] Show that, given any basis of M, the matrix A of T with respect to that basis satisfies $\det A = \det T$.

(b) Deduce the multiplicativity of determinants from the above: $\det(TS) = (\det T)(\det S)$. Many of the properties of determinants can be similarly obtained from the alternating property and multilinearity.

(c) Now suppose $1 \leqslant n \leqslant r$. Consider $\Lambda^n T : \Lambda^n M \to \Lambda^n M$. If T has matrix A with respect to a basis $e_1, \ldots, e_r$ of M, and $e_{i_1} \wedge \cdots \wedge e_{i_n}$ and $e_{j_1}^\vee \wedge \cdots \wedge e_{j_n}^\vee$ are basis elements of $\Lambda^n M$ and $(\Lambda^n M)^\vee \cong \Lambda^n M^\vee$ as in (x) above, show that the matrix entry of $\Lambda^n T$ corresponding to $e_{i_1} \wedge \cdots \wedge e_{i_n}$ and $e_{j_1}^\vee \wedge \cdots \wedge e_{j_n}^\vee$, namely

$$\langle e_{j_1}^\vee \wedge \cdots \wedge e_{j_n}^\vee, \Lambda^n T(e_{i_1} \wedge \cdots \wedge e_{i_n}) \rangle,$$

equals the $n \times n$-minor of A corresponding to $(j_1, \ldots, j_n; i_1, \ldots, i_n)$. Thus, the $n \times n$ minors of a linear transformation T have coordinate-free interpretation in terms of $\Lambda^n T$, as the various $\langle v^\vee, \Lambda^n T(v) \rangle$, where $v^\vee$ varies along $M^\vee$ and v along M.

(d) Now the the characteristic polynomial of an R-linear endomorphism $T \in \mathrm{End}_R(M)$ of a free R-module M of finite rank has been defined independently of coordinates, since the determinant has been. Show that the characteristic polynomial of T can also be described as:

$$\sum_{i=0}^{n} (-1)^i \operatorname{tr}(\Lambda^i T) x^{n-i}.$$

In other words, the coefficients of the characteristic polynomial of T are, up to signs, traces of various exterior powers of T.

Hint: One approach is to prove this for diagonalizable matrices over $\mathbb{C}$, extend by density to $\mathrm{M}_n(\mathbb{C})$, deduce for $\mathbb{Z}[x_{ij} | 1 \leqslant i, j \leqslant n]$, and then use this to extend to general R: this strategy has been discussed in Lecture 2 while discussing one of the proof of the Cayley-Hamilton theorem.

[23] This lets us define $\det T$ without using coordinates. Without this, one can choose a basis of M, let A be the matrix of T with respect to that basis, and *define* $\det T$ to be $\det A$: since it would still be independent of the choice of the basis, as $\det(BAB^{-1}) = \det A$, and hence well-defined. But it is still more elegant, and often practically useful, to have its direct definition without recourse to bases.

8.7. **Some applications of tensor, symmetric and exterior powers.** Let $\mathbb{P}^n(\mathbb{C})$ denote the set of lines in $\mathbb{C}^n$ through the origin. More generally, given a vector space V over $\mathbb{C}$, let $\mathbb{P}(V)$ denote the set of one-dimensional subspaces of V. In algebraic geometry, $\mathbb{P}(V)$ is realized as an algebraic variety.

Example 8.23.

9. Lecture 9 — various kinds of bilinear forms, and quadratic forms

Throughout today's lecture, R is a commutative ring, where we assume $1 \neq 0$ to be safe. For any R-module M, we set $M^\vee = \mathrm{Hom}_R(M, R)$. Unless otherwise stated, M will be an R-module. We say "M is finite free over R" to mean M is a free R-module of finite rank. Later, we will specialize to the case where $R = F$ is a field.

9.1. Symmetric, skewsymmetric and alternating bilinear forms.

Definition 9.1. (i) A bilinear form (with respect to R) on an R-module M is a bilinear map $B : M \times M \to R$, i.e., an element of $\mathrm{Bil}_R(M) := \mathrm{Bil}_R(M, M; R) \cong (T^2M)^\vee$.

(ii) We denote by $B \mapsto {}^tB$ the "swapping" map $\mathrm{Bil}_R(M) \to \mathrm{Bil}_R(M)$, i.e., given by action of the nontrivial element of $\mathfrak{S}_2$ on $\mathrm{Bil}_R(M, M; R) = (T^2M)^\vee$. Thus, ${}^tB(m, n) = B(n, m)$ for all $m, n \in M$.

(iii) A bilinear form $B : M \times M \to R$ is called
- symmetric, if ${}^tB = B$, i.e., if it factors through $T^2M \to S^2(M)$;
- alternating, if $B(m, m) = 0$ for all $m \in M$, i.e., if B factors through $T^2(M) \to \Lambda^2(M)$;
- skew-symmetric, if ${}^tB = -B$, i.e., if $B(m, n) = -B(n, m)$ for all $m, n \in M$.

This gives subsets $\mathrm{SymBil}_R(M), \mathrm{AltBil}_R(M), \mathrm{SkewsymBil}_R(M) \subset \mathrm{Bil}_R(M)$. The subscript R may be dropped if it is understood.

Remark 9.2. (i) As a special case of our discussion on Hom-tensor adjointness, we have:

$$\begin{array}{ccccc} \mathrm{Hom}_R(M, M^\vee) & \xleftarrow{B \mapsto B(m,-)} & \mathrm{Bil}_R(M) & \xrightarrow{B \mapsto B(-,m)} & \mathrm{Hom}_R(M, M^\vee)\ . \\ & & \uparrow{\scriptstyle -\circ(M\times M \to T^2M)} & & \\ & & \mathrm{Hom}_R(T^2M, R) = (T^2M)^\vee & & \end{array}$$

Thus, when M is free of finite rank, $\mathrm{Bil}_R(M) \cong (T^2M)^\vee$ can also be described as $M^\vee \otimes_R M^\vee = T^2(M^\vee)$, using the pairing between $M^{\otimes n}$ and $(M^\vee)^{\otimes n}$ from Lecture 8. This also lets us view a bilinear form on M as a "noncommutative degree two homogeneous polynomial on M".

(ii) Like with tensors, any alternating form is skew-symmetric, and the converse is true when $2 \in R^\times$ but not in general. When $2 \in R^\times$, we will also see later that symmetric bilinear forms are essentially the same as quadratic forms (to be defined later today). The objects of primary interest seem to be quadratic and alternating forms, rather than symmetric and skewsymmetric forms (except that the latter are essentially the former when $2 \in R^\times$).

(iii) We will mostly only be interested in bilinear forms on M when M is a free R-module of finite rank, but this notion is typically interesting at least when M is a projective R-module of finite rank, a notion that we have not defined.

(iv) When $2 = 0$ in R, skewsymmetric is the same as symmetric, but as mentioned above, these notions are of secondary interest in this case.

(v) It follows from the definition that $\mathrm{SymBil}(M), \mathrm{AltBil}(M), \mathrm{SkewsymBil}(M) \subset \mathrm{Bil}(M)$ identify with

$$(S^2M)^\vee = ((T^2M)^\vee)^{\mathfrak{S}_2} \subset (T^2M)^\vee, \quad (\Lambda^2 M)^\vee \subset (T^2M)^\vee \quad \text{and } ((T^2M)^\vee)^{\mathfrak{S}_2,\mathrm{sgn}},$$

respectively. [24] For some more descriptions, see the following exercise.

Exercise 9.3. Hopefully understanding the following can clarify the context.

(i) Let a group G act on an R-module M, and let $\chi : G \to R^\times$ be a character. Recall, from HW 4, the submodule $M^{(G,\chi)} \subset M$ of (G,χ)-invariants of M, and the quotient module $M_{(G,\chi)}$ of (G,χ)-coinvariants of M. G acts on $M^\vee := \mathrm{Hom}_R(M,R)$ by $(g \cdot f)(m) = f(g^{-1} \cdot m)$.
   (a) The quotient map $M \to M_{(G,\chi)}$ induces an injection $(M_{G,\chi})^\vee \hookrightarrow M^\vee$. Show that this injection defines an isomorphism $(M_{(G,\chi)})^\vee \to (M^\vee)^{G,\chi^{-1}}$.
   (b) In contrast, we also have a map $(M^\vee)_{G,\chi^{-1}} \to (M^{G,\chi})^\vee$, because the restriction map $M^\vee \to (M^{G,\chi})^\vee$ factors through $M^\vee \to (M^\vee)_{G,\chi^{-1}}$, but this may not be an isomorphism.
   (c) Conclude from (a) that when M is free of finite rank, $(S^nM)^\vee$ identifies with the space of symmetric tensors in $T^nM^\vee$ and $(\Lambda^n M)^\vee$ identifies with the space of antisymmetric tensors in $T^nM^\vee$.
   Hint: By an exercise from Lecture 8, since M is free of finite rank, $(T^nM)^\vee$ identifies with $T^nM^\vee$ $\mathfrak{S}_n$-equivariantly (i.e., the pairing of T^nM with $T^nM^\vee$ from Lecture 8 satisfies $\langle g \cdot \lambda, g \cdot \mu\rangle = \langle \lambda, \mu\rangle$ for all $g \in \mathfrak{S}_n$ – verify this). Be warned that S^nM doesn't always identify with the space of symmetric tensors in M.

(ii) Assume that M is a finite free R-module. Define a sequence of "obvious maps"

$$\mathrm{SymBil}(M) \cong (S^2M)^\vee \cong ((T^2M)^\vee)^{\mathfrak{S}_2} \cong (T^2M^\vee)^{\mathfrak{S}_2} = \{\text{Symmetric tensors in } T^2M^\vee\} \to S^2(M^\vee),$$

where the last arrow is an isomorphism when $2 \in R^\times$ but not in general (more about this in the discussion on quadratic forms).

(iii) *(To be double-checked)* Assume that M is a finite free R-module. Define a sequence of "obvious maps"

$$\mathrm{AltBil}(M) \cong (\Lambda^2M)^\vee \cong ((T^2M)^\vee)^{\mathfrak{S}_2,\mathrm{sgn}} \cong (T^2M^\vee)^{\mathfrak{S}_2,\mathrm{sgn}} = \{\text{Antisymmetric tensors in } T^2M^\vee\} \to \Lambda^2(M^\vee),$$

where these maps are all isomorphisms, even when $2 \notin R^\times$.
Hint/note: The last map in the above line is not

$$\{\text{Antisymmetric tensors in } T^2M^\vee\} \hookrightarrow T^2M^\vee \to \Lambda^2 M^\vee,$$

since this is not an isomorphism in general. Rather, the proof that $\Lambda^2 M$ was free (for free finite rank M) gave us an injection $\Lambda^2 M \to T^2(M)$, whose image was the space of alternating tensors in M. The last map is inverse to this, with M replaced by $M^\vee$. This also gives the isomorphism $(\Lambda^2M)^\vee \cong \Lambda^2(M^\vee)$ from HW 4.

[24] This is true regardless of R, because care has been applied in deciding the order of the decorations: e.g., S^2M is a *quotient module* of T^2M, so $(S^2M)^\vee$ is a *submodule* of $(T^2M)^\vee$.

Lemma 9.4. *Assume that $2 \in R^\times$. Then the inclusions of $(S^2M)^\vee, (\Lambda^2 M)^\vee$ in $(T^2M)^\vee$ induce an isomorphism $(T^2M)^\vee \cong (S^2M)^\vee \oplus (\Lambda^2 M)^\vee$. In other words, when $2 \in R^\times$, every bilinear form is uniquely a sum of a symmetric bilinear form and a skew-symmetric (or equivalently since $2 \in R^\times$, an alternating) bilinear form.*

Proof. Easy; use $B = (B + {}^tB)/2 + (B - {}^tB)/2$. □

9.2. **Matrices associated to bilinear forms, and determinant.**

Exercise 9.5. Let $M = R^n$. Write $e_1, \dots, e_n$ for the standard basis of R^n. Show that there is an R-module isomorphism

$$\mathrm{Bil}_R(M) \to M_n(R),$$

sending $B \in \mathrm{Bil}_R(M)$ to $[B(e_i, e_j)]_{1 \leqslant i,j \leqslant n} \in \mathrm{M}_n(R)$, and whose inverse sends $A \in M_n(R)$ to

$$(X, Y) \mapsto {}^tXAY,$$

with $X, Y \in R^n$ thought of as column vectors in the obvious way.

Definition 9.6. Let M be a free finite rank R-module. If $B \in \mathrm{Bil}_R(M)$, then for any basis $e_1, \dots, e_n$ of M, the matrix of B with respect to that basis is defined to be

$$A = [B(e_i, e_j)]_{1 \leqslant i,j \leqslant n}.$$

Remark 9.7. (i) In the context of Definition 9.6, note that the basis $e_1, \dots, e_n$ identifies M with R^n. Either by transporting B to R^n via this isomorphism and using Exercise 9.5, or (better) directly, note that we can describe $B \in \mathrm{Bil}_R(M)$ in terms of its image $A \in \mathrm{M}_n(R)$ as:

$$B\Big(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\Big) = {}^tXAY, \tag{39}$$

where X, Y are the column vectors defined by $(x_1, \dots, x_n) \in R^n$ and $(y_1, \dots, y_n) \in R^n$. It easily follows that sending B to its matrix with respect to $e_1, \dots, e_n$ defines an R-module isomorphism (depending on the choice of $e_1, \dots, e_n$)

$$\mathrm{Bil}_R(M) \to \mathrm{M}_n(R). \tag{40}$$

(ii) Let $M \cong R^n$. If $A \in \mathrm{M}_n(R)$ is the matrix of $B \in \mathrm{Bil}_R(M)$ with respect to some basis, then regardless of this choice of basis, $B \in \mathrm{SymBil}_R(M)$ if and only if A is a symmetric matrix, $B \in \mathrm{AltBil}_R(M)$ if and only if $A = [a_{ij}]$ is an alternating matrix in the sense that $a_{ij} = -a_{ji}$ and $a_{ii} = 0$ for all $1 \leqslant i, j \leqslant n$, and $B \in \mathrm{SkewsymBil}_R(M)$ if and only if $A = [a_{ij}]$ is a skewsymmetric matrix.

(iii) While the isomorphism $\mathrm{Bil}_R(M) \to \mathrm{M}_n(R)$ of (40) depends on the choice of the basis $e_1, \dots, e_n$, this dependence is easy to describe. Namely, if $e'_1, \dots, e'_n$ is another basis of M, say $e'_i = \sum_k p_{ik} e_k$ for each $1 \leqslant i \leqslant n$ for some matrix $P = [p_{ij}]_{1 \leqslant i,j \leqslant n} \in$

$\mathrm{M}_n(R)$, then automatically $P \in \mathrm{GL}_n(R)$, and the matrix of B with respect to $e'_1, \dots, e'_n$ is given by $A' = [a'_{ij}]$, where

$$a'_{ij} = P(e'_i, e'_j) = B\Big(\sum_k p_{ik} e_k, \sum_l p_{jl} e_l\Big) = \sum_{k,l} p_{ik} B(e_k, e_l) p_{jl} = \sum_{k,l} p_{ik} a_{kl} p_{jl} = (P A\, {}^t P)_{i,j},$$

so that $A' = P \cdot A \cdot {}^t P$. Thus, the matrix of B with respect to $e'_1, \dots, e'_n$ is $P \cdot A \cdot {}^t P$, with $P \in \mathrm{GL}_n(R)$ relating the e_i to the e'_i as above.

Exercise 9.8. Let M be a finite free R-module. If the matrix of $B \in \mathrm{Bil}_R(M)$ with respect to $e_1, \dots, e_n$ is A, then with respect to the choices of $e_1, \dots, e_n$ as a basis for M and the dual basis $e_1^\vee, \dots, e_n^\vee$ as a basis for $M^\vee$, show that the matrix of the linear transformations $m \mapsto B(m, -)$ and $m \mapsto B(-, m)$ are ${}^t A$ and A.

Definition 9.9. (i) Let M be a finite free R-module. The determinant $\det B$ of $B \in \mathrm{Bil}_R(M)$ is an element of $R/R^{\times 2}$, the quotient of the set R under the multiplicative action of the group $R^{\times 2}$ (i.e., the set [25] of $R^{\times 2}$-orbits on R), defined in either of the following equivalent ways:

(a) *Definition using coordinates/matrices.* $\det B$ is defined to be the image of $\det A$ in $R/R^{\times 2}$, where A is the matrix of B with respect to any basis of M: this is well-defined, because replacing the basis will replace A by some $P \cdot A \cdot {}^t P$, with $P \in \mathrm{GL}_n(R)$ (see Remark 9.7(iii) above), and $\det(P \cdot A \cdot {}^t P) = (\det A)(\det P)^2 \in (R^\times)^2 \cdot \det A$ (recall/note that elements of $\mathrm{GL}_n(R)$ have determinant in $R^\times$).

(b) *Coordinate-free definition.*

- If M is free of rank one, then $B(x, x)$, where x is any generator of M, has a well-defined image in $R/R^{\times 2}$, which we call $\det B$: any other generator is of the form ax with $a \in R^\times$, and $B(ax, ax) = a^2 B(x, x) \in B(x, x) \cdot R^{\times 2}$.
- For general M that is free of finite rank n, a problem from HW4 associates to B a bilinear form $\Lambda^n B \in \mathrm{Bil}_R(\Lambda^n M)$. Since $\Lambda^n M$ is of rank one, the rank one case gives a definition of $\det(\Lambda^n B)$, and we set $\det B = \det(\Lambda^n B)$.

To see that these definitions agree, look back at the definitions, and make sense (in $R/R^{\times 2}$) of:

$$\det B \overset{\text{second notion}}{=} \Lambda^n B(e_1 \wedge \cdots \wedge e_n, e_1 \wedge \cdots \wedge e_n) = \det([B(e_i, e_j)]_{1 \leqslant i,j \leqslant n}) \overset{\text{first notion}}{=} \det B.$$

(ii) Recall that associated to B are two maps $M \to M^\vee$, namely $m \mapsto B(m, -)$ and $m \mapsto B(-, m)$.

- B is said to be nondegenerate if these maps are both injective: i.e., given $m \in M$, there exist $m', m'' \in M$ such that $B(m, m') \neq 0 \neq B(m'', m)$.
- B is said to be perfect if one of these maps is an isomorphism. Easy exercise: this is equivalent to both of these maps being isomorphisms (e.g., use Remark 9.10(i) below).

[25] it is just a set and not a group or anything.

(iii) Let M and M' be R-modules, $B \in \mathrm{Bil}_R(M)$ and $B' \in \mathrm{Bil}_R(M')$. The direct sum of the pairs (M, B) and (M', B'), denoted $(M, B) \oplus (M', B')$ or $(M, B) \perp (M', B')$, is defined to be $(M \oplus M', B \oplus B')$, where $B \oplus B'$ is the bilinear form on $M \oplus M'$ such that for all $m_1, m_2 \in M$ and $m_1', m_2' \in M'$ we have:

$$(B \oplus B')((m_1, m_1'), (m_2, m_2')) = B(m_1, m_2) + B'(m_1', m_2').$$

Remark 9.10. (i) It is immediate that either of $m \mapsto B(m-)$ or $m \mapsto B(-, m)$ is an isomorphism if and only if $\det B \in R/R^{\times 2}$ belongs to $R^\times/R^{\times 2}$: use Exercise 9.8.

(ii) If B is perfect, then it is nondegenerate, but not conversely (e.g., $B : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ given by $B(m, n) = 2mn$ is nondegenerate, but not perfect). However, note that if R is a field, then since M and $M^\vee$ are vector spaces of the same dimension, nondegenerate B are perfect as well. Thus, once we specialize to fields, we will simply write nondegenerate, but "perfect" will also be understood.

(iii) Henceforth, we will be interested in pairs (M, B) consisting of a finite free R-module M, and $B \in \mathrm{Bil}_R(F)$. It is clear how to define isomorphisms between such pairs (M, B) and (M', B'): they should be R-module isomorphisms $T : M \to M'$ that pull B' back to B (i.e., $B'(Tm_1, Tm_2) = B(m_1, m_2)$ for all $m_1, m_2 \in M$). Isomorphisms between such pairs will also be referred to as 'isometries'. In contrast, I haven't noticed much interest in the notion of just morphisms (as opposed to isomorphisms) between such pairs.

Exercise 9.11. (i) Verify any claim in Remark 9.10 that you don't see immediately.

(ii) Show that isometries preserve the various notions you have seen above, such as determinant, nondegenerate and perfect.

(iii) Let M, M' be finite free R-modules. If bilinear forms $B \in \mathrm{Bil}_R(M)$ and $B' \in \mathrm{Bil}_R(M')$ have matrices A and A' with respect to choices of bases $e_1, \ldots, e_n$ of M and $e_1', \ldots, e_{n'}'$ of M', then the bilinear form $B \oplus B' \in \mathrm{Bil}_R(M \oplus M')$ has matrix, with respect to the basis $e_1, \ldots, e_n, e_1', \ldots, e_n'$, matrix of the form

$$\begin{pmatrix} A & \\ & A' \end{pmatrix}$$

(where the entries in the 'blank' blocks are understood to be all 0).

9.3. **Quadratic forms.**

Definition 9.12. Let M be a finite free R-module.

(i) A quadratic form on M is a function $q : M \to R$ that satisfies any of the following four equivalent conditions, their equivalence following from Exercise 9.13(i) below:

(a) It is given by a homogeneous polynomial of degree 2: i.e., $\exists f = \sum l_i l_i' \in S^2(M^\vee)$, with each $l_i, l_i' \in M^\vee$, such that for all $m \in M$ we have $f(m) = \sum l_i(m) l_i'(m)$.

(b) We have $q(am) = a^2 q(m)$ for all $a \in R$ and $m \in M$, and the map $B_q : M \times M \to R$ given by $B_q(m, n) := q(m + n) - q(m) - q(n)$ is an R-bilinear form on M. We will call B_q the bilinear form associated to q.

(c) There exists $B \in \mathrm{Bil}_R(M)$ such that for all $m \in M$, we have $q(m) = B(m,m)$. In this case, we might write $q = q_B$, and call q the quadratic form associated to B.

(d) Using some [26] basis of M to identify it with R^n, there exist $a_i \in R$ for $1 \leqslant i \leqslant n$ and $a_{ij} \in R$ for $1 \leqslant i < j \leqslant n$, such that for all $(x_1, \ldots, x_n) \in R^n = M$, we have $q(x_1, \ldots, x_n) = \sum_{1 \leqslant i \leqslant n} a_i x_i^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij} x_i x_j$.

In short, just "a homogeneous polynomial of degree 2".

(ii) Let $\mathrm{Quad}(M) = \mathrm{Quad}_R(M)$ denote the R-module of quadratic forms on M. The fact that each element of $\mathrm{Quad}(M)$ is defined by an element of $S^2(M^\vee)$ gives us a surjective R-module homomorphism $S^2(M^\vee) \to \mathrm{Quad}(M)$, which is an isomorphism by Exercise 9.13(ii) below.

(iii) By a quadratic space over R we will refer to a pair (M, q), where M is a free R-module of finite rank, and $q \in \mathrm{Quad}_R(M)$ is a quadratic form. It is clear how to define isomorphisms of quadratic spaces, and those will be referred to as isometries.

Exercise 9.13. (i) Show the equivalence of the four definitions of quadratic forms.
Hint: It could be easier to relate the condition (d) to the remaining three conditions. Anyway, please do make sure you can do this exercise.

(ii) Show that the R-module homomorphism $S^2(M^\vee) \to \mathrm{Quad}(M)$ is an isomorphism.
Hint: Without loss of generality, let $M = R^n$. By the previous exercise, assume $q(x_1, \ldots, x_n) = \sum_{1 \leqslant i \leqslant j \leqslant n} a_{ij} x_i x_j$. If some $a_{ii} \neq 0$, take $m = e_i$. If each $a_{ii} = 0$ and some $a_{ij} \neq 0$, take $m = e_i + e_j$.

(iii) Let M be a free R-module of finite rank. Show that we have a commutative diagram with exact rows, each of whose columns are isomorphisms:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathrm{AltBil}_R(M) & \xrightarrow{\text{incl.}} & \mathrm{Bil}_R(M) & \xrightarrow{B \mapsto q_B} & \mathrm{Quad}_R(M) & \longrightarrow & 0\ . \\
 & & \downarrow & & \downarrow & & \downarrow{\scriptstyle \text{(ii)}} & & \\
0 & \longrightarrow & \Lambda^2(M^\vee) & \longrightarrow & T^2(M^\vee) & \longrightarrow & S^2(M^\vee) & \longrightarrow & 0
\end{array}$$

Again, $\Lambda^2(M^\vee) \to T^2(M^\vee)$ is the not-so-obvious inclusion, since $\Lambda^2(M^\vee)$ is a priori a quotient of $T^2(M^\vee)$: one is using the exercise from Lecture 8 proving that $\Lambda^2(M^\vee)$ is a free R-module (when M is finite free).

(iv) Recall B_q and q_B from (b) and (c) of Definition 9.12(i). Show that the composite of $q \mapsto B_q$ and $B \mapsto q_B$, in either direction, is not the identity, but multiplication by 2. Thus, if $2 \in R^\times$, these maps are not "lossy" at all, [27] and in this case, it may be more convenient to replace B_q by

$$B'_q := (1/2)B_q : (x, y) \mapsto 2^{-1} \cdot (q(x+y) - q(x) - q(y)),$$

which ensures $q_{B'_q} = q$ and $B'_{q_B} = B$.

[26] or as is easily seen to be equivalent, any.

[27] which they are if, say $2 = 0$ in R.

(v) Verify that one way to transcribe the above exercise is as follows: When $2 \in R^\times$, so that $\mathrm{Bil}_R(M) = \mathrm{SymBil}_R(M) \oplus \mathrm{AltBil}_R(M)$, the surjection $\mathrm{Bil}_R(M) \to \mathrm{Quad}_R(M)$ given by $B \mapsto q_B$ vanishes on $\mathrm{AltBil}_R(M)$, and restricts to an isomorphism $\mathrm{SymBil}_R(M) \to \mathrm{Quad}_R(M)$. This isomorphism also respects various constructs we will define in what follows, such as the radical etc.

(vi) On the other hand, when $2 \notin R^\times$, $\mathrm{SymBil}_R(M) \hookrightarrow \mathrm{Bil}_R(M) \to \mathrm{Quad}_R(M)$ fails quite spectacularly to be an isomorphism; say when $2 = 0$ in R (e.g., if $R = \mathbb{F}_2$):
- Any element of $\mathrm{AltBil}_R(M) \subset \mathrm{Bil}_R(M)$ is contained in $\mathrm{SymBil}_R(M)$, but maps to 0 in $\mathrm{Quad}_R(M)$: thus, these give B such that $q_B = 0$. Specifically, consider $M = R^2$, $B((x_1, y_1), (x_2, y_2)) = x_1y_2 + x_2y_1$: B is nondegenerate, but $q_B = 0$.
- While every element of $\mathrm{Quad}_R(M)$ is still of the form q_B for some $B \in \mathrm{Bil}_R(M)$, there may not exist such $B \in \mathrm{SymBil}_R(M)$; e.g., $M = R^2$, $Q(x, y) = xy$.

We may see these more explicitly in examples later.

9.4. **Sesquilinear forms.** Recall that complex inner products, being only conjugate-linear in one of the variables, are not bilinear. So what generalizes complex inner products are not bilinear forms, but sesquilinear forms ("1.5-linear forms"), to define which we need a 'conjugation', so that we can talk of conjugate-linearity.

In this subsection, we will take $R = F$ to be a field, and let E/F be a separable quadratic extension. We will probably not use the following facts from field theory, but it is probably good to keep them in mind/make sure you can prove them:

(i) If $\mathrm{char}\, F \neq 2$, then $E = F[\alpha] = F[\sqrt{a}]$, for some nonsquare $a \in F$, with $\alpha^2 = a$.

(ii) If $\mathrm{char}\, F = 2$, then $E = F[\alpha]$, for some root α of a quadratic polynomial of the form $x^2 - x - a = 0$ (if α satisfies $x(x - b) + c = 0$, then $b \neq 0$ by separability, so replace α by αb, which gives the same extension).

It is easy to see that $\mathrm{Gal}(E/F) := \mathrm{Aut}_{F\text{-}Alg}(E) = \{1, \sigma\}$, where σ swaps the roots α and $-\alpha$ of $x^2 - a = 0$ (if $\mathrm{char}\, F \neq 2$) or the roots α and $1 - \alpha$ of $x^2 - x - a = 0$ (if $\mathrm{char}\, F = 2$).

Notation 9.14. (i) Given E/F and $1 \neq \sigma \in \mathrm{Gal}(E/F)$ as above, we will typically denote σ by $a \mapsto \bar{a}$.

(ii) We will write $\mathrm{tr}_{E/F} : E \to F$ and $N_{E/F} : E^\times \to F^\times$ for the trace and norm maps associated to E/F, which we recall are defined by: $\mathrm{tr}_{E/F}(a) = a + \bar{a} \in F$ and $N_{E/F}(a) = a\bar{a} \in F$. Alternatively, $a \mapsto (b \mapsto ab)$ defines a ring homomorphism $E \to \mathrm{End}_F(E)$, whose composite with $\mathrm{tr} : \mathrm{End}_F(E) \to F$ and $\det : \mathrm{End}_F(E) \to F$ define $\mathrm{tr}_{E/F}$ and $N_{E/F}$, respectively (see Exercise 9.18 below).

Definition 9.15. Fix a separable quadratic extension E/F as above.

(i) Let V be a vector space over E. An E/F-sesquilinear form on V is a $\mathbb{Z}$-bilinear map

$$B : V \times V \to E,$$

such that for all $v, w \in V$ and $a \in E$, we have:

$$B(av, w) = \bar{a}B(v, w), \qquad \text{and } B(v, aw) = aB(v, w).$$

Thus, a sesquilinear form is linear in the second variable and "conjugate-linear" in the first variable. "Sesqui" means 1.5.

(ii) Sesquilinear forms on V form an E-vector space of dimension equal to $(\dim V)^2$, which will be denoted by $\mathrm{SesLin}_{E/F}(V) = \mathrm{SesLin}(V)$.

(iii) It is immediately checked that we have a linear involution $* : \mathrm{SesLin}(V) \to \mathrm{SesLin}(V)$, denoted $B \mapsto B^*$, such that $B^*(v,w) = \overline{B(w,v)}$ for each $v, w \in V$: note that B^* is indeed sesquilinear, unlike $(v,w) \mapsto B(w,v)$.

(iv) A sesquilinear form B is called Hermitian if $B^* = B$, and skew-Hermitian if $B^* = -B$. This gives us subsets $\mathrm{Herm}_{E/F}(V) = \mathrm{Herm}(V)$, $\mathrm{SkewHerm}_{E/F}(V) = \mathrm{SkewHerm}(V) \subset \mathrm{SesLin}(V)$, which are not E-subspaces (if B is Hermitian or skew-Hermitian, then for $a \in E \backslash F$, aB will usually not be Hermitian or skew-Hermitian): they are F-subspaces of $\mathrm{SesLin}(V)$.

Notation 9.16. To study E/F-sesquilinear forms on an E-vector space V, it is helpful to consider the E-vector space $\bar{V}$ whose underlying abelian group is V, on which E operates through σ followed by the usual scalar multiplication: its elements can be denoted $\{\bar{v} \mid v \in V\}$, where each $\bar{v}$ is a formal symbol, and its vector space structure defined by $a\bar{v} + \bar{w} = \overline{\bar{a}v + w}$.

Remark 9.17. Let E^σ be the E-algebra $\sigma : E \to E$, thought of as an (E,E)-bimodule. It is an easy exercise to see that that $\bar{V}$ has an obvious identification with the E-vector space $E^\sigma \otimes_E V$. We will probably not use this today, but this sort of consideration is quite common in mathematics.

Exercise 9.18. Use your knowledge of tensor products to show that the two definitions of $\mathrm{tr}_{E/F}$ and $N_{E/F}$ coincide.

Hint: It is enough to show that the eigenvalues of $b \mapsto ab$ are a and $\sigma(a)$, respectively. To see this, it is natural to consider $E \otimes_E E$, the extension of scalars of the F-vector space E to an E-vector space. Show that, as an E-vector space, we have an isomorphism $E \otimes_F E \cong E \oplus E$ of E-vector spaces, which transports "multiplication by a" to the E-linear transformation of $E \oplus E$ that is "multiplication by $\bar{a}$" on the first factor and "multiplication by a" on the second.

Exercise 9.19. In these questions, unless otherwise stated, E/F is a separable quadratic extension, and V is an E-vector space.

(i) Show that $B \mapsto (v \mapsto B(v,-))$ and $B \mapsto (v \mapsto B(-,v))$ define E-linear isomorphisms $\mathrm{SesLin}(V) \to \mathrm{Hom}_E(\bar{V}, V^\vee)$ and $\mathrm{SesLin}(V) \to \mathrm{Hom}_E(V, \bar{V}^\vee)$. Here we identify the underlying sets of V and $\bar{V}$ via $v \mapsto \bar{v}$.

(ii) For any E-basis $e_1, \ldots, e_n$ of V, define the matrix of B with respect to this basis to be $A = [B(e_i, e_j)]_{1 \leqslant i,j \leqslant n}$. Give an "$X^*AX$" description for B, where $X^* = \sigma({}^tX)$, σ being applied entry-wise to ${}^tX$. Similarly, give a "PAP^*" change of basis formula for this matrix.

(iii) For any choice of basis of E, show that $B \in \mathrm{SesLin}(V)$ is Hermitian if and only if its matrix A is Hermitian (i.e., $A^* = A$), and that B is skew-Hermitian if and only if its matrix A is skew-Hermitian (i.e., $A^* = -A$).

(iv) For $B \in \mathrm{SesLin}(V)$, define what $\det B$ should be, as an element of $E/N_{E/F}(E^\times)$. Define both coordinates-based and coordinate-free variants of this notion, and prove their equivalence.

(v) Define the notion of nondegeneracy for $B \in \mathrm{SesLin}(V)$, which also agrees with what perfectness should be, and show that B is nondegenerate (i.e., perfect) if and only if $\det B$ is not 0 inside $E/N_{E/F}(E^\times)$.

(vi) Let $B \in \mathrm{SesLin}(V)$. Show that if B is Hermitian, then $\det B \in F$. Show also that if B is skew-Hermitian and $\dim V$ is odd (resp., even), then $\det B \in \ker(\mathrm{tr}_{E/F}) = \{x \in E \mid \mathrm{tr}_{E/F}(x) = 0\}$ (resp., $\det B \in F$). Note that $\ker(\mathrm{tr}_{E/F})$ is a one-dimensional F-subspace of E just like F is.
Hint: Use that $\det A^* = \overline{\det A}$.

(vii) Define direct sums and isometries for pairs (V, B) consisting of an E-vector space V and an E/F-sesquilinear form $B : V \times V \to E$.

(viii) Unlike the difference between symmetric and skew-symmetric forms, the difference between Hermitian and skew-Hermitian forms is not serious: if $0 \neq \beta \in E$ is such that $\mathrm{tr}_{E/F}(\beta) = 0$ (if $\mathrm{char}\, F \neq 2$, one can take $\beta \in E \backslash F$ such that $\beta^2 \in F$, and otherwise one can take $\beta = 1$), multiplication by β gives an F-linear isomorphism $\mathrm{Herm}(V) \to \mathrm{SkewHerm}(V)$.

Remark 9.20. In the theory of algebraic groups, orthogonal groups are defined as groups of isometries of (nondegenerate) quadratic forms, symplectic groups as groups of isometries of (nondegenerate) alternating forms, and unitary groups as groups of isometries of (nondegenerate) Hermitian (sesquilinear) forms. By Exercise 9.19(viii), unitary groups are also groups of isometries of skew-Hermitian (sesquilinear) forms.

9.5. **Radicals, orthogonals etc.** Henceforth, we will only consider the case when $R = F$ is a field, or when we have a quadratic separable extension E/F of fields.

Definition 9.21. Consider one of the following scenarios:

(a) V is a vector space over a field F, and B is a symmetric or alternating bilinear form on V. We will only use the 'symmetric' case when $\mathrm{char}\, F \neq 2$.

(b) E/F is a quadratic separable field extension, V is a vector space over E, and $B \in \mathrm{Herm}(V)$ or $B \in \mathrm{SkewHerm}(V)$.

Then:

(i) For $S \subset V$, we define the orthogonal of S with respect to B to be:

$$S^\perp := \{v \in V \mid B(v, S) = \{0\}\} = \{v \in V \mid B(S, v) = \{0\}\}$$

(the two descriptions agree since we are in the symmetric/alternating/Hermitian/skew-Hermitian situation). Note that $S^\perp \subset V$ is a subspace (i.e., an F-subspace in the bilinear case, and an E-subspace in the sesquilinear case).

(ii) The radical of B, denoted $\mathrm{rad}(B)$, is defined to be $V^\perp$, so $\mathrm{rad}(B) = \{v \in V \mid B(v, V) = \{0\}\} = \{v \in V \mid B(V, v) = \{0\}\}$.
V is nondegenerate (or equivalently, perfect) if and only if $\mathrm{rad}(B) = 0$ (easy).

Definition 9.22. Let (V, q) be a quadratic space over a field F. If $\operatorname{char} F \neq 2$, we define the radical $\operatorname{rad}(q)$ of q to be $\operatorname{rad}(B_q)$, and say that q is nondegenerate/perfect if B_q is. However, a definition that works independently of its characteristic is given as follows:

(i) *Radical.* The radical of (V, q) is defined to be

$$\operatorname{rad}(q) := \{v \in \operatorname{rad}(B_q) \mid q(v) = 0\},$$

which is readily checked to be a vector subspace of V (If $q(v) = q(w) = 0$ with $v, w \in \operatorname{rad}(B_q)$, then $q(v + w) = q(v) + q(w) + B_q(v, w) = 0$).

(ii) *Nondegeneracy.*
   (a) (V, q) is regular if $\operatorname{rad}(q) = 0$.
   (b) (V, q) is nondegenerate if one of the following three conditions, which can be shown to be equivalent, are satisfied, where q_K is obtained from q by extending scalars along $F \hookrightarrow K$ (exercise: make this precise):
      - $\operatorname{rad}(q_K) = 0$ for every field K containing F;
      - $\operatorname{rad}(q_K) = 0$ for some algebraically closed field K containing F; and
      - (V, q) is regular and $\dim_F \operatorname{rad}(B_q) \leqslant 1$.

Exercise 9.23. If $\operatorname{char} F \neq 2$, show that $\operatorname{rad}(q) = \operatorname{rad}(B_q)$, since $q|_{\operatorname{rad}(B_q)}$ can be recovered from $B_q|_{\operatorname{rad}(B_q)}$ (easy/immediate). Note that when $\operatorname{char} F = 2$, taking $M = F$ and $q(x) = x^2$, $\operatorname{rad}(B_q) = F$ but $\operatorname{rad}(q) = 0$. Hopefully this gives at least a very partial explanation of why we defined 'radical' and 'nondegenerate' for quadratic forms the way we did.

Remark 9.24. The point of giving uniform characteristic-independent definitions as above (which we did not discuss in the lecture) is that, with these uniform definitions, a lot of the nontrivial properties of quadratic forms and orthogonal groups outside characteristic two also work over characteristic two. I haven't justified this, and am appealing to your faith here: the appeal-to-faith point is that often when one identifies the 'correct' general definitions, a lot of things extend to the general case, but identifying the 'correct' general definitions, or even realizing that such exist, might involve work.

Exercise 9.25.
(i) Let (V, B) be as in Definition 9.21. Show that B descends to a bilinear form $\bar{B}$ on $V/\operatorname{rad}(B)$, and that $\operatorname{rad}(\bar{B}) = 0$.

(ii) We do not have a canonical maximal nondegenerate subspace of (V, B), rather we have several noncanonical ones: show that any subspace $W \subset V$ that is complementary to $\operatorname{rad}(B)$ satisfies that $(W, B|_W)$ is nondegenerate. (Easy). Moreover, $V \to V/\operatorname{rad}(B)$ restricts to an isometry from $(W, B|_W)$ to $(V/\operatorname{rad}(B), \bar{B})$.

(iii) Formulate and prove the analogue of the above questions for quadratic spaces (V, q).

Example 9.26.
(i) *The case where* $\dim_F V = 1$ *or* $\dim_E V = 1$*, as appropriate*:

- Bilinear cases: For $V = F$, $\operatorname{Bil}_F(F) = \operatorname{SymBil}_F(F) = \{(B_a : (x, y) \mapsto axy) \mid a \in F\}$, and $\operatorname{AltBil}_F(F) = 0$. Note that $\det B_a = aF^{\times 2}$. It is easy to see that B_a and B_b are isometric if and only if $a \in bF^{\times 2}$, i.e., $\det B_a = \det B_b$.
  It easily follows that sending (V, B) to $\det B$ gives a bijection

$$\{\text{Isometry classes of } (V, B) \text{ with } \dim_F V = 1 \text{ and } B \in \operatorname{Bil}_F(V)\} \to F/F^{\times 2},$$

and we can replace $\mathrm{Bil}_F(V)$ with $\mathrm{SymBil}_F(V)$ in the above assertion. All the (V, B) are symmetric in this case, none alternating. (F, B_a) is a representative with determinant $aF^{\times 2}$.

- Sesquilinear cases: For E/F quadratic separable and $V = E$, $\mathrm{SesLin}_{E/F}(E) = \{(B_a : (x, y) \mapsto a\bar{x}y \mid a \in E\}$, $\mathrm{Herm}_{E/F}(E) = \{B_a \mid a \in F\}$, $\mathrm{SkewHerm}_{E/F}(E) = \{B_a \mid a \in \ker(\mathrm{tr}_{E/F})\}$. We have $\det(B_a) = aN_{E/F}(E^\times)$, and B_a is isometric to B_b if and only if $\det B_a = \det B_b$.
  It easily follows that sending (V, B) to $\det B$ gives a bijection

$$\{\text{Isometry classes of } (V, B) \text{ with } \dim_E V = 1 \text{ and } B \in \mathrm{SesLin}_{E/F}(V)\} \to E/N_{E/F}(E^\times).$$

  We have a similar assertion involving $\mathrm{Herm}_{E/F}(V)$ and $F/N_{E/F}(E^\times)$, and one involving $\mathrm{SkewHerm}_{E/F}(V)$ and $\ker(\mathrm{tr}_{E/F})/N_{E/F}(E^\times)$.
- Quadratic cases: Sending V to $q(v)F^{\times 2}$, where $0 \neq v \in V$ is any basis vector, gives a bijection

$$\{\text{Isometry classes of } (V, q) \text{ with } \dim_F V = 1\} \to F/N_{E/F}(E^\times).$$

  A representative with determinant (the image of) a is given by (F, ax^2).

(ii) *Hyperbolic planes.*

(i) In the context of symmetric or Hermitian (resp., alternating or skew-Hermitian) forms, a hyperbolic plane over F or E is a pair (V, B) that has, with respect to some basis, matrix

$$\begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \quad \left(\text{resp., } \begin{pmatrix} & -1 \\ 1 & \end{pmatrix}\right).$$

(ii) In the context of quadratic spaces, a hyperbolic plane over F is a quadratic space that is isomorphic to $(F^2, (x, y) \mapsto xy)$.

Note that every hyperbolic plane is nondegenerate. A hyperbolic space is defined to be a direct sum of hyperbolic planes.

(iii) In all the five contexts, a vector $0 \neq v \in V$ is called isotropic if $B(v, v) = 0$ or $q(v) = 0$ as appropriate, and anisotropic otherwise.

(iv) A subspace $W \subset V$ is called anisotropic if every $0 \neq w \in W$ is anisotropic, isotropic if it is not anisotropic, and totally isotropic if $B|_W = 0$ or $q|_W = 0$, as appropriate.

10. Lecture 10 — various kinds of bilinear forms, and quadratic forms (contd.)

Notation 10.1. In this lecture, (V, B) will denote a finite dimensional vector space together with a symmetric, alternating, Hermitian or skew-Hermitian form as above, associated to a field F or a separable quadratic extension E/F as appropriate. A subspace $W \subset V$ will refer to an F-subspace in the former two cases, and to an E-subspace in the latter two cases. We will often omit to mention the finite dimensionality of various pairs (V, B) as above, but it should be considered as understood.

Further, in the symmetric case, we will assume $\operatorname{char} F \neq 2$. In the bilinear cases (i.e., in the symmetric and the alternating cases), if we say E we will mean F, if we say $\bar{V}$, we will mean V, etc.

We recall that when $\operatorname{char} F = 2$, symmetric is the same as skew-symmetric, but neither of these notions works well. However, quadratic and alternating forms do work well in characteristic 2.

If you don't like this level of generality, at least on a first reading just think of the symmetric bilinear case. Recall the notion of direct sums of bilinear forms, nondegenerate, orthogonal, radical etc.

10.1. **Building (V, B) from smaller subspaces.** In many situations we can get direct sum decompositions, thanks to the following simple lemma:

Lemma 10.2. *(i) Suppose (V, B) is as in Notation 10.1. If a subspace $W \subset V$ is nondegenerate, then $V = W \oplus W^{\perp}$ (i.e., $(V, B) = (W, B|_W) \oplus (W^{\perp}, B|_{W^{\perp}})$.*
(ii) If instead V is nondegenerate, we still have $\dim W + \dim W^{\perp} = \dim V$.

Remark 10.3. In (ii) of the lemma, we may not have $V = W \oplus W^{\perp}$, since $W \cap W^{\perp}$ may be nonzero: e.g., W can be a totally isotropic subspace of V, in which case $W \subset W^{\perp}$. For an even more particular example, suppose (V, B) is a hyperbolic plane, where $V = Fe_1 + Fe_2$ with $B(e_1, e_1) = B(e_2, e_2) = 0$ and $B(e_1, e_2) = 1$. Then if W equals Fe_1 or Fe_2, we have $W = W^{\perp}$.

Proof of Lemma 10.2. For ease of reading, we will discuss only the symmetric bilinear case: for the Hermitian and the skew-Hermitian cases, simply put a "$\bar{}$" at the appropriate places. For both assertions, by the rank-nullity theorem, it is enough to show that the following sequence is exact, where we recall that $W^{\vee} : \operatorname{Hom}_F(W, F)$:

$$0 \to W^{\perp} \hookrightarrow V \overset{v \mapsto B(v,-)}{\to} W^{\vee} \to 0$$

(this would give $\dim W + \dim W^{\perp} = \dim W^{\vee} + \dim W^{\perp} = \dim V$; in the situation of (i), since $W \cap W^{\perp} = 0$, this would also give $V = W \oplus W^{\perp}$).

The exactness at $W^{\perp}$ and V are immediate, while for the surjectivity of $V \to W^{\vee}$:

- In the situation of (i), use that $B|_W$ is perfect (nondegenerate), so already the restriction $W \to W^\vee$ of $V \to W^\vee$, given by $v \mapsto B(v, -)$, is surjective.
- In the situation of (ii), $V \to W^\vee$ is the composite of the surjective map $V^\vee \to W^\vee$ (given by restriction to W) and the isomorphism $V \to V^\vee$ given by $v \mapsto B(v, -)$.

□

Exercise 10.4. This exercise is trivial, but is helpful to keep in mind; it will be used without further comment in what follows: if (V, B) is one of the four kinds of pairs above, and is nondegenerate, and we have a decomposition $(V, B) \cong (W_1, B_1) \oplus (W_2, B_2)$, then (W_1, B_1) and (W_2, B_2) are nondegenerate.

By Lemma 10.2, to decompose (V, B), it is enough to construct smaller nondegenerate subspaces within (V, B). Here are two ways to do this:

- Any anisotropic vector spans such a subspace.
- Any isotropic vector in V is contained in a hyperbolic plane in V, by the following lemma.

Lemma 10.5. *Let (V, B) be as in Notation 10.1 (thus, $\operatorname{char} F \neq 2$ in the symmetric bilinear case), and assume that (V, B) is nondegenerate. If $0 \neq v \in V$ is an isotropic vector, then (V, B) contains a hyperbolic plane (for its type) containing v.*

Proof. We have $B(v, v) = 0$. It is enough to find $w \in V$ such that $B(w, w) = 0 \neq B(v, w)$: for, then we can scale w to ensure $B(v, w) = 1$, and then $\operatorname{Span}(v, w)$ will be a hyperbolic plane.

Since (V, B) is nondegenerate, there exists $w' \in V$ such that $B(v, w') \neq 0$. It is enough to find a scalar a such that $B(w' - av, w' - av) = 0$: then $w := w' - av$ satisfies $B(v, w) = B(v, w') \neq 0 = B(w, w)$. Thus, we need $B(w', w') = B(w', av) + B(av, w')$, which can be ensured case-by-case:

- In the alternating case, this is automatic.
- In the symmetric case (where $\operatorname{char} F \neq 2$ by assumption), choose $a = B(v, w')/2$.
- In the Hermitian case, choose $a \in E$ such that $\operatorname{tr}_{E/F}(aB(w', v)) = B(w', w') \in F$ (such an a exists since $\operatorname{tr}_{E/F} : E \to F$ is surjective, because E/F is separable).
- In the skew-Hermitian case, choose $a \in E$ such that $aB(w', v) - \overline{aB(w', v)} = B(w, w)$ (choose c such that $\operatorname{tr}_{E/F} c = 0$, and then a such that $\operatorname{tr}_{E/F}(caB(w', v)) = cB(w', w') \in F$ (explanation for why $cB(w', w') \in F$: $\bar{c} = -c$, $\overline{B(w', w')} = -B(w', w')$, so $\overline{cB(w', w')} = cB(w', w')$, so $cB(w', w') \in F$): this is the same trick we used in Lecture 9 to say that skew-Hermitian is not essentially different from Hermitian.

□

Corollary 10.6. *In the alternating cases, we have (of course noncanonically)*

$$(V, B) = \left(\bigoplus_{i=1}^{n} \mathbb{H} \right) \oplus (W, 0),$$

where $\mathbb{H}$ stands for an alternating hyperbolic plane, and $(W, 0)$ is the radical of B.

Proof. Choosing any complement W' of $\mathrm{rad}(B)$ to write V as a direct sum of $(W, 0)$ and a nondegenerate subspace $(W', V|_{W'})$ (this was an exercise, Exercise 9.25((ii)), from Lecture 9), we reduce to the nondegenerate case. Now, any nonzero vector in V is isotropic, so we can inductively apply Lemma 10.5 together with Lemma 10.2(i) ($\mathbb{H}$ is nondegenerate; we also use the easy Exercise 10.4, which will typically not even be cited henceforth). □

Remark 10.7. Thus, the classification of alternating forms is uniform over all fields (apparently, also over local rings and Dedekind domains; see Professor Nair's notes).

Corollary 10.8. *Let (V, B) belong to the symmetric, Hermitian or skew-Hermitian cases. Then there is a basis of V for which the matrix of B is diagonal: there exists a basis $e_1, \ldots, e_n$ of V and nonzero $a_1, \ldots, a_r$ belonging to E or F, where $r = n - \dim \mathrm{rad}(B) \leqslant n$, such that*

$$B(e_i, e_j) = \begin{cases} 0, & \text{if } i \neq j \text{ or } i = j > r, \text{ and} \\ a_i, & \text{if } 1 \leqslant i = j \leqslant r. \end{cases}$$

Proof. As in Corollary 10.6, we choose a complement of $\mathrm{rad}(B)$ to reduce to the case where (V, B) is nondegenerate. We can induct using Lemma 10.2(i), if show that any nondegenerate (V, B) has an anisotropic vector. This is Lemma 10.9 below. □

Lemma 10.9. *Any nondegenerate symmetric bilinear, Hermitian or skew-Hermitian space (V, B) has an anisotropic vector.*

Proof. Let $v, w \in V$ with $B(v, w) \neq 0$. If $B(v, v) \neq 0$ or $B(w, w) \neq 0$, we are done. Otherwise:

- In the symmetric case, if v and w are isotropic, then $B(v+w, v+w) = 2B(v, w) \neq 0$.
- In the Hermitian case, choose $a \in E$ so that $\mathrm{tr}_{E/F}(aB(v, w)) \neq 0$; then if v and w are anisotropic, $B(v + aw, v + aw) = \mathrm{tr}_{E/F}(aB(v, w)) \neq 0$.
- The skew-Hermitian case is similar, with $\mathrm{tr}_{E/F}(aB(v, w))$ replaced by $aB(v, w) - \overline{aB(v, w)}$ (see the proof of Lemma 10.5).

□

Henceforth, we will mostly ignore the skew-Hermitian case: as seen in Lecture 9 and above, we can reduce the proofs in the skew-Hermitian case to those in the Hermitian case, using a trace zero element in E.

Corollary 10.10. *Let (V, B) belong to the symmetric or Hermitian cases. Let $\mathcal{E}$ be a set of representatives for $F/F^{\times 2}$ in the symmetric case, and a set of representatives for $F/N_{E/F}(E^\times)$ in the Hermitian case.*

(i) V has a basis relative to which B has a matrix of the form $\mathrm{diag}(a_1, \ldots, a_n)$, where each a_i belongs to $\mathcal{E}$.

(ii) Assume that we are in the symmetric case (so $\operatorname{char} F \neq 2$*) and that* F *is algebraically closed. Sending* (V, B) *to* $(\dim V, \dim \operatorname{rad}(B))$ *gives a bijection*

$$\{\textit{Isometry classes of pairs } (V,B) \textit{ over } F \textit{ with } B \textit{ symmetric bilinear}\} \to \{(n,m) \in \mathbb{N}^2 \mid m \leqslant n\}.$$

(iii) (Sylvester's law of inertia) Assume that we are either in the symmetric case with $F = \mathbb{R}$*, or in the sesquilinear case with* $E/F = \mathbb{C}/\mathbb{R}$*. Then there is a bijection*

$$\{\textit{Isometry classes of pairs } (V,B) \textit{ in this setting}\} \to \{(p,q,r) \in \mathbb{N}^3\},$$

sending (V, B) *to* (p, q, r) *where* $r = \dim \operatorname{rad}(B)$*, and* p *(resp.,* q*) is the maximal possible dimension of a* B*-positive definite (resp.,* B*-negative definite) subspace of* V*.*

Proof. Let us see (i). Corollary 10.8 gives us a basis $e_1, \dots, e_n$ of V, relative to which B has matrix of the form $\operatorname{diag}(a_1, \dots, a_n)$. Choose nonzero c_i such that $a_i c_i^2 \in \mathcal{E}$ (in the symmetric case) or $a_i N_{E/F}(c_i) \in \mathcal{E}$ (in the Hermitian case). Then, relative to the basis $c_1 e_1, \dots, c_n e_n$ of V, the matrix of B is diagonal, with entries in $\mathcal{E}$.

If F is algebraically closed, and (V, B) is symmetric bilinear, we can take $\mathcal{E} = \{0, 1\}$, and (ii) follows.

Now we come to (iii). In both the symmetric and the Hermitian cases, we can take $\mathcal{E} = \{0, 1, -1\}$. Thus, every (V, B) in this setting has a basis with respect to which its matrix is of the form $\operatorname{diag}(1^p, -1^q, 0^r)$, where we have written 1^p for a sequence of p 1's, etc. It is clear that if $e_1, \dots, e_n$ is such a basis, then $r = \dim \operatorname{rad}(B)$, B is positive definite on the p-dimensional subspace spanned by $e_1, \dots, e_p$, and it is negative definite on the q-dimensional subspace spanned by $e_{p+1}, \dots, e_{p+q}$.

To check that p is the maximal dimension of a positive definite subspace of V – and hence uniquely determined – note that no positive definite subspace of V can intersect $\operatorname{Span}(e_{p+1}, \dots, e_n)$, and hence has dimension at most p. Similarly, q is the maximal dimension of a negative definite subspace of V, and is hence uniquely determined. □

Definition 10.11. If (V, B) is a symmetric bilinear, Hermitian or skew-Hermitian form, then a basis $e_1, \dots, e_n$ of V is said to be an orthogonal basis (with respect to B) if its elements are pairwise B-orthogonal, i.e., if $B(e_i, e_j) = 0$ for $i \neq j$, i.e., if B has a diagonal matrix with respect to this basis.

Remark 10.12. By Corollary 10.10, every quadratic form q on a vector space V over a field F with $\operatorname{char} F \neq 2$ can be diagonalized, i.e., there exists a basis $e_1, \dots, e_n$ of V and $a_1, \dots, a_n \in F$ such that $q(\sum_{i=1}^n x_i e_i) = \sum_{i=1}^n a_i x_i^2$, for all $x_1, \dots, x_n \in F$. However, this is not true when F has characteristic two: note that when $\operatorname{char} F = 2$, if q can be diagonalized, then $B_q = 0$, but we have several q such that $B_q \neq 0$, e.g., a hyperbolic plane q given by $q(x_1 e_1 + x_2 e_2) = x_1 x_2$.

Exercise 10.13. Above, we saw the classification of symmetric nondegenerate bilinear forms over the real numbers. Read up about the classification of symmetric nondegenerate bilinear forms/quadratic forms over finite fields, at least outside characteristic two. The

main result in this context is: if (V, B) and (V', B') are nondegenerate, then they are isometric if and only if $\dim V = \dim V'$ and $\det B = \det B'$. Since $\#(\mathbb{F}_q^\times / \mathbb{F}_q^{\times 2}) = 2$, it follows that there are exactly two isometry classes of nondegenerate quadratic forms over $\mathbb{F}_q$ of a given dimension, when q is odd.

Example 10.14. A special case of Sylvester's law of inertia helps us understand why there are three cases of nondegenerate conics: ellipses, parabolas and hyperbolas. Recall that a conic in $\mathbb{R}^2$ is the subset given by a degree two polynomial $f(x, y) = f_2(x, y) + f_1(x, y) + f_0(x, y)$, where $f_2(x, y) = ax^2 + bxy + cy^2$ is a homogeneous quadratic polynomial, f_1 is a linear polynomial, and f_0 is a constant. Denote this conic by C_f.

Thus, f_2 is a quadratic form. If this quadratic form is positive definite or negative definite (i.e., has signature $(2, 0, 0)$ or $(0, 2, 0)$, then C_f is an ellipse. If it has signature $(1, -1, 0)$ or $(-1, 1, 0)$, then C_f is a hyperbola. Otherwise f_2 is degenerate, and C_f is either a parabola or 'degenerate' in the sense of being a line or a product of two lines, etc. Thus, finding the shape of C_f basically amounts to diagonalizing the quadratic form f_2.

In higher dimensions, i.e., in $\mathbb{R}^n$, too, one considers solutions of quadratic equations in n variables. Such solutions are called quadrics (as opposed to conics, which are in $\mathbb{R}^2$). It is clear that Sylvester's law of inertia helps us study and classify quadrics in $\mathbb{R}^n$ too: e.g., in $\mathbb{R}^3$, one gets objects such as elliptic hyperboloids, parabolic hyperboloids etc., distinguished from each other by the signature of the associated quadratic form (namely, the "homogeneous of degree two part" of the given equation).

Corollary 10.15. *Let (V, B) belong to the symmetric, Hermitian or skew-Hermitian cases. Then we have a decomposition*

$$(V, B) = (V_h, B_h) \oplus (V_a, B_a) \oplus (\operatorname{rad} B, 0),$$

where (V_h, B_h) is a direct sum of hyperbolic planes, and (V_a, B_a) is anisotropic.

Proof. As before, reduce to the case where (V, B) is nondegenerate. If (V, B) is anisotropic, we are done. If not, we have a hyperbolic plane in (V, B) by Lemma 10.5, so we can use Lemma 10.2 and induct. □

Remark 10.16. In the above corollary, the subspaces (V_h, B_h) and (V_a, B_a) of (V, B) are not uniquely determined. It so turns out, nevertheless, that their isometry classes are uniquely determined: this is not obvious, and will be proved in Corollary 10.22 as a consequence of Witt's theorem.

The following slight strengthening of Lemma 10.5 will help us reduce the proof of Witt's theorem to the nondegenerate case.

Lemma 10.17. *Let (V, B) belong to the symmetric, Hermitian or skew-Hermitian cases. Assume that (V, B) is nondegenerate.*

(i) Given any totally isotropic subspace $W_0 \subset V$, with basis $e_1, \ldots, e_s$, there exists a totally isotropic subspace $U_0 \subset V$, and a basis $f_1, \ldots, f_s$ of U_0, such that $B(e_i, f_j) = \delta_{i,j}$ for $1 \leqslant i, j \leqslant s$ (then automatically, $W_0 + U_0 = W_0 \oplus U_0$ inside V).

(ii) Let $W \subset V$ be an arbitrary subspace. Write $W = W_0 \oplus W_1$, where $W_0 = \mathrm{rad}(B|_W)$, and W_1 is any complement to W_0 in W (thus, $B|_{W_1}$ is nondegenerate). Let $e_1, \dots, e_s$ be a basis for W_0. Then there exists a totally isotropic subspace $U_0 \subset W_1^\perp \subset V$, and a basis $f_1, \dots, f_s$ of U_0, such that $B(e_i, f_j) = \delta_{i,j}$ for $1 \leqslant i, j \leqslant s$ (thus, $W_0 + W_1 + U_0 = W_0 \oplus W_1 \oplus U_0$ inside V).

Proof. (ii) follows from applying (i) to $W_1^\perp \supset W_0$ (which is nondegenerate as V and W_1 are), so it is enough to prove (i). By nondegeneracy, which implies that $v \mapsto B(v, -)$ defines a bijection $V \to V^\vee$, we can choose $f_1' \in V$ such that $B(e_i, f_1') = \delta_{i,1}$ for all $1 \leqslant i \leqslant s$. As in the proof of Lemma 10.5, there exists a scalar a such that $B(f_1' - ae_1, f_1' - ae_1) = 0$. Setting $f_1 = f_1' - ae_1$, it is still true that $B(e_i, f_1) = \delta_{i,1}$ for all $1 \leqslant i \leqslant s$ (use that W_0 is totally isotropic). $V_1 := \mathrm{Span}(e_1, f_1)$ is a hyperbolic plane, and is hence nondegenerate, and $e_2, \dots, e_s$ belong to $V_1^\perp$. Now we can induct. □

(i) of the lemma above can be summarized as follows: given a totally isotropic subspace $W_0 \subset V$, there exists a totally isotropic subspace $U_0 \subset V$ such that B restricts to a perfect pairing between W_0 and U_0. Automatically, $W_0 + U_0 = W_0 \oplus U_0 \subset V$ is a direct sum of hyperbolic planes. The configuration of (ii) of the lemma, with $W_0 \oplus W_1 \oplus U_0 \subset V$, and with B restricting to a nondegenerate form on W_1 and to a perfect pairing between the totally isotropic subspaces W_0 and U_0 of V, is also something very commonly seen. Writing W_1 in between W_0 and U_0 is for reasons to do with group theory (algebraic groups).

10.2. **Witt's theorem: statement and consequences.** Recall that whenever we talk of a symmetric bilinear form, we assume that $\mathrm{char}\, F \neq 2$.

Theorem 10.18 (Witt's theorem/Witt's extension theorem)**.** *Let (V, B) be a nondegenerate symmetric bilinear, alternating bilinear, Hermitian or skew-Hermitian form, and let $h : (W, B|_W) \to (W', B|_{W'})$ be an isometry between subspaces $W, W' \subset V$. Then there exists $g \in \mathrm{Aut}(V, B)$ such that $g|_W = h$. In other words, h can be extended to an isometry $V \to V$.*

Here is another way to state the same theorem:

Theorem 10.19 (Witt's theorem, slight restatement)**.** *Let (V, B) and (V', B') be isometric nondegenerate symmetric, alternating, Hermitian or skew-Hermitian spaces, and let $h : (W, B|_W) \to (W', B|_{W'})$ be an isometry between subspaces $W \subset V, W' \subset V'$. Then there exists an isometry $g : (V, B) \to (V', B')$ such that $g|_W = h$.*

Theorem 10.19 is a formal consequence of Theorem 10.18 (please make sure you understand that this is trivial), so only the former will be proved. Before proving Theorem 10.18, let us derive some corollaries.

Corollary 10.20 (Witt cancellation theorem)**.** *Let (V, B) and (V', B') be nondegenerate symmetric, alternating, Hermitian or skew-Hermitian spaces. Assume that*

$$(V, B) = (W_1, B_1) \oplus (W_2, B_2), \qquad \text{and} \qquad (V', B') = (W_1', B_1') \oplus (W_2', B_2').$$

If $(V, B) \cong (V', B')$ and $(W_1, B_1) \cong (W_1', B_1')$, then $(W_2, B_2) \cong (W_2', B_2')$ (where "$\cong$" means "is isometric to").

Proof, assuming Witt's theorem. Let $h : W_1 \to W_1'$ be any isometry, and use Theorem 10.19 to extend it to an isometry $g : V \to V'$. Then automatically, g takes $W_1^\perp$ to $W_1'^\perp$ (every isometry behaves well with respect to taking '$\perp$').

It follows from the nondegeneracy of V that that W_1, W_2, W_1', W_2' are nondegenerate (by now, it shouldn't be necessary to quote Exercise 10.4 here), and (using Lemma 10.2) that the inclusions $W_2 \subset W_1^\perp$ and $W_2' \subset W_1'^\perp$ are equalities, so g takes $(W_2, B|_{W_2}) = (W_2, B_2)$ to $(W_2', B'|_{W_2'}) = (W_2', B_2')$. □

Corollary 10.21. *Let (V, B) be as above, nondegenerate. All maximal totally isotropic subspaces of (V, B) are $\mathrm{Aut}(V, B)$-translates of each other, and hence have the same dimension.*

Proof, assuming Witt's theorem. If $W, W' \subset V$ are maximal totally isotropic and $\dim W \leqslant \dim W'$, then any injection $h : W \hookrightarrow W'$ is an isometry, and hence extends to an isometry $g : V \to V$. Then $g^{-1}(W')$ is a totally isotropic subspace containing W, and hence equals W by maximality. □

Corollary 10.22. *Let (V, B) be as above, nondegenerate.*

(i) All maximal hyperbolic subspaces of (V, B) are $\mathrm{Aut}(V, B)$-translates of each other, and hence have the same dimension.

(ii) Write, using Corollary 10.15:

$$(V, B) \cong \mathbb{H}^s \oplus (V_{an}, B_{an}),$$

where $\mathbb{H}^s$ is an s-fold sum of hyperbolic planes $\mathbb{H}$, and (V_{an}, B_{an}) is anisotropic. Then s and the isometry class of (V_{an}, B_{an}) are uniquely determined by (V, B).

Proof, assuming Witt's theorem. (i) follows exactly as in Corollary 10.21.

Let us prove (ii). Given a decomposition $(V, B) = (W_1, B_1) \oplus (W_2, B_2)$, we claim that $W_1 \subset V$ is a maximal hyperbolic subspace if and only if (W_2, B_2) is anisotropic: to see "$\Rightarrow$", use Lemma 10.5; to see "$\Leftarrow$", if $W_1' \supsetneq W_1$ is a hyperbolic subspace, and hence a direct sum of its own totally isotropic subspaces, then the image of W_1' under the projection $V \to W_2$ has a nonzero isotropic vector, a contradiction. Thus, any $\mathbb{H}^s \hookrightarrow V$ as in (ii) is a maximal hyperbolic subspace of V, and s is half its dimension, so that the uniqueness of s follows from (i). The uniqueness of the isometry class of (V_{an}, B_{an}) then follows from the Witt cancellation theorem (Corollary 10.20). Note that in the alternating case, of course, $s = \dim V/2$, and $V_{an} = 0$. □

Definition 10.23. In the context of the above corollary, (the isometry class of) (V_{an}, B_{an}) is called the anisotropic kernel of (V, B).

10.3. **The proof of Witt's theorem in the symmetric bilinear case.** We will only prove Theorem 10.18 in the symmetric bilinear case.

Step 1. Reduction to the case where $W \subset V$ is nondegenerate. Let $W_0 = \mathrm{rad}(B|_W) \subset W$ have basis $e_1, \dots, e_s$. Consider the basis $e'_1 := h(e_1), \dots, e'_s := h(e_s)$ of $W'_0 := h(W_0) = \mathrm{rad}(B|_{W'})$. Let W_1 be any complement to W_0 in W, so that $W'_1 := h(W_1)$ is a complement to W'_0 in W'.

Associate to W, W_0, W_1 and $e_1, \dots, e_s$, the subspace U_0 and its basis $f_1, \dots, f_s$ as in Lemma 10.17(ii). Associate, similarly, U'_0 and $f'_1, \dots, f'_s$ to W', W'_0, W'_1 and $e'_1, \dots, e'_s$.

We extend h to a map $W + U_0 \to W_1 + U_0$, still denoted h, as follows:

$$h : W + U_0 = W_0 \oplus W_1 \oplus U_0 \to W'_0 \oplus W'_1 \oplus U'_0 = W' + U'_0,$$

that agrees with h on $W = W_0 \oplus W_1$ (and in particular takes each e_i to e'_i), and such that $h(f_i) = f'_i$ for each i. It is immediately verified (please do verify) that this new h is an isometry on $W_0 + U_0 = W_0 \oplus U_0$ as well as on W_1, and hence on $W_0 + W_1 + U_0 = W + U_0$. It is enough to extend this new h to g.

Since B is nondegenerate on each of W_1 and $W_0 + U_0$, it is nondegenerate on $W + U_0$ (it is an easy exercise to prove that a direct sum of nondegenerate subspaces is nondegenerate). Similarly, it is nondegenerate on $W' + U'_0$. Thus, if we know how to extend isometries from nondegenerate $W \subset V$, the general case follows.

Remark 10.24. To go ahead, the following very obvious principles will be helpful, and we will use them repeatedly (so please make sure you are very clear with their justifications):

- If $(V, B) = (W, B|_W) \oplus (W^\perp, B|_{W^\perp})$, then any isometry $g_0 : W \to W$ has a unique extension to an isometry $g : V \to V$ that acts as the identity on $W^\perp$.
- If $(V, B) = (W, B|_W) \oplus (W^\perp, B|_{W^\perp})$ is nondegenerate, then any isometry $g : V \to V$ that maps W into itself also maps $W^\perp$ into itself.

Step 2. Reduction to smaller subspaces. The induction will sort of be on $\dim W$, rather than on $\dim V$. The induction is carried out by the following lemma.

Lemma 10.25. *Assume that $W_0, W_1 \subset V$ are nondegenerate, and orthogonal to each other. Suppose:*

- *Any isometry $h_0 : W_0 \hookrightarrow V$ extends to an isometry $g_0 : V \to V$; and*
- *Any isometry $h_1 : W_1 \hookrightarrow W_0^\perp \supset W_1$ extends to an isometry $g_1 : W_0^\perp \to W_0^\perp$, and hence to an isometry $g_1 : V \to V$ that acts as the identity on W_0 (see Remark 10.24; we won't repeat this henceforth).*

Then any isometry $h : W_0 \oplus W_1 \to V$ extends to an isometry $g : V \to V$.

Proof. Write $W = W_0 \oplus W_1$. Extend $h_0 := h|_{W_0}$ to an isometry $g_0 : V \to V$. Note that $g_0^{-1}h : W \hookrightarrow V$ is an isometry, and we have $(g_0^{-1}h)|_{W_0} = h_0^{-1}h_0 =$ identity, so that $g_0^{-1}h$ maps $W_1 \subset W_0^\perp$ into $W_0^\perp$. Therefore, we can extend $h_1 := g_0^{-1}h|_{W_1}$ to an isometry $g_1 : V \to V$ that acts as the identity on W_0.

Consider $g = g_0 g_1$: this is clearly an isometry of V, and it restricts to h on W because:

- On W_0, it equals $g_0 = h_0$.
- On W_1, it equals $g_0 h_1 = g_0 \cdot g_0^{-1} h|_{W_1} = h|_{W_1}$.

□

One proof of Witt's extension theorem that we will describe in the symmetric bilinear case, will use the notion of reflections, that are interesting and important in their own right.

Definition 10.26. Let V be a vector space over F, and B a symmetric nondegenerate bilinear form on it. Let $v \in V$ be an anisotropic vector, so that $V = Fv \oplus v^\perp$ ($v^\perp \subset V$ is the hyperplane orthogonal to v). Then the reflection r_v associated to v, or the reflection about the hyperplane orthogonal to v, is the unique isomorphism $V \to V$ that sends v to $-v$, and fixes $v^\perp$.

Clearly, any reflection $r_v : V \to V$ is an isometry. Check that, concretely, it can be given by the following formula (it is enough to check it separately on v and on $v^\perp$):

$$r_v(w) = w - \frac{2B(v,w)}{B(v,v)} v.$$

Let us emphasize that we have defined reflections only in the symmetric bilinear case. There are somewhat related notions in the other cases, but those behave differently, and seem to be less useful. Note also that we associate reflections only to anisotropic vectors.

Proof of Witt's theorem. The symplectic case is much easier, and is left as an exercise. The skew-Hermitian case follows from the Hermitian case, since these two cases are not "essentially different", by an observation from Lecture 9. Thus, we consider the symmetric and the Hermitian cases.

Since V is nondegenerate, we know that there is an anisotropic vector $v \in W$ (Lemma 10.9). By the previous lemma and induction, we are reduced to proving the following (after possibly replacing V with a smaller subspace):
Step 3. The lemma is true when $W = Fv$ for some anisotropic vector $v \in V$. Thus, assume that $W = Fv$.

Let $h(v) = w$, so $B(v,v) = B(w,w)$. It is enough to show that there exists an isometry $g : V \to V$ with $g(v) = w$. For this we give two arguments.

An argument that applies only in the symmetric case. This argument follows Professor Nair's notes, which follows Scharlau's book.

If v and w are parallel, then $w = \pm v$, so we can take g to be either the reflection r_v or multiplication by ± 1, both of which are isometries. Thus, assume that v and w span a two-dimensional subspace of v.

Case 1. $v-w$ is anisotropic. In this case, there exists a reflection r_{v-w} about the hyperplane orthogonal to $v - w$. Since $B(v,v) = B(w,w)$, $v + w$ is orthogonal to $v - w$, and hence

$r_{v-w}(v+w) = v+w$, while $r_{v-w}(v-w) = -(v-w)$. This forces $r_{v-w}(v) = w$, as desired (note that we have used the invertibility of 2 here).

Case 2. $v-w$ is isotropic. Since $B(v-w, v-w)+B(v+w, v+w) = 2B(v,v)+2B(w,w) = 4B(v,v) \neq 0$ (recall that $2 \neq 0$ in F by assumption), we get that $-v-w$ is anisotropic. Thus, the argument of Case 1 applies with v replaced by $-v$, and gives that $r_{-v-w} = r_{v+w}$ sends $-v$ to w. Thus, $r_{v+w}r_v$ sends v to w, as desired.

This finishes the symmetric case, now we consider a more general one.

An argument that applies in the symmteric and the Hermitian cases (not discussed in the lecture). This argument follows Serge Lang's book.

If v is parallel to w, say $v = aw$ with a a scalar, then since $B(v,v) = B(w,w)$, we have $a = \pm 1$ (in the symmetric bilinear case) or $N_{E/F}(a) = 1$ (in the Hermitian case). Therefore, we can take g to be multiplication by a, which is in either case an isometry. Assume therefore that v, w span a two-dimensional subspace of V.

Case 1. $\mathrm{Span}(v,w) \subset V$ is a nondegenerate subspace. If $B(v,w) \neq 0$, then the unique scalar a such that $B(v,w) = aB(w,v)$ satisfies the following property: $a = 1$ in the symmetric case, and $N_{E/F}(a) = 1$ in the Hermitian case (where $B(v,w) = \overline{B(w,v)}$). If $B(v,w) = 0$, choose any a with this property, and we automatically have $B(v,w) = aB(w,v)$. Since $\mathrm{Span}(v,w)$ is nondegenerate, we get a self-isometry of $\mathrm{Span}(v,w)$, extending $h : Fv \to V$, by sending v to w and w to av (that it is an isometry needs to be checked only on pairs involving the basis elements v and w, which is easy in this case). This can be extended to a self-isometry $g : V \to V$ that acts as the identity on $(\mathrm{Span}(v,w))^{\perp}$ (see Remark 10.24). Clearly, g extends h, as desired.

Case 2. $\mathrm{Span}(v,w)$ is a degenerate subspace. Let v_0 span $\mathrm{rad}(\mathrm{Span}(v,w))$ (which cannot be the whole of $\mathrm{Span}(v,w)$, since v is anisotropic). Let $w = av + bv_0$, with a, b scalars. Since $B(v,v) = B(w,w)$, we get $a = \pm 1$ or $N_{E/F}(a) = 1$. We can replace v by av (since multiplication by a is an isometry), so we may now assume that $w = v + bv_0$. Scaling v_0 if necessary, we have $w = v + v_0$. By Lemma 10.17(ii), there exists $u_0 \in v^{\perp}$ such that $B(v_0, u_0) = 1$, so $B(u_0, v_0) = 1$ as well. $\mathrm{Span}(u_0, v, v_0) = \mathrm{Span}(u_0, w, v_0)$ is nondegenerate.

Set $a = -B(v,v)^{-1}$ (we are reusing the letter a: the earlier a was different and serves no purpose now). We claim that there exists a scalar b such that $b + \bar{b} = -a\bar{a}B(v,v) = -\overline{B(v,v)}^{-1}$: here, we write $\bar{b}$ for b itself in the symmetric case, and for its $\mathrm{Gal}(E/F)$-conjugate in the Hermitian case. In the symmetric case, this follows from the hypothesis that $2 \in F^{\times}$, while in the Hermitian case, this is the case because $\mathrm{tr}_{E/F}$ is surjective. One then checks that the following defines an isometry of $\mathrm{Span}(u_0, v, v_0)$:

$$v_0 \mapsto v_0, v \mapsto v + v_0 = w, u_0 \mapsto u_0 + av + bv_0$$

(check on each pair of basis elements). This isometry sends v to w. Since $\mathrm{Span}(u_0, v, v_0)$ is nondegenerate, this isometry can be extended to V (Remark 10.24). $\square$

The above proof has the following corollary:

Corollary 10.27. *Let (V, B) be a nondegenerate symmetric bilinear form over F (as usual, $\operatorname{char} F \neq 2$), and let $\sigma : W \to W'$ be an isometry between two subspaces of V. Then there is a product of at most $2 \cdot \dim V$ reflections in B that restricts to σ. In particular, every element of the orthogonal group $O(V, B) := \operatorname{Aut}(V, B)$ is a product of at most $2(\dim V)$-many reflections. If further (V, B) is anisotropic, then we need at most $\dim V$-many reflections.*

Proof of Corollary 10.27. Easy exercise going through the proof of Theorem 10.18. Consider the crucial step where $W = Fv$ and we had to find an isometry $V \to V$ taking v to w. If $v = w$ nothing needed to be done; if $v = -w$ one can use r_v; when $v - w$ was anisotropic, this was accomplished with one reflection r_{v-w}; and otherwise it was accomplished with 2 reflections r_v and r_{v+w}. □

Remark 10.28.
(i) Something stronger than Corollary 10.27 holds: in its setting, every element of $O(V, B)$ is a product of $\dim V$-many reflections, just like in the anisotropic case. This is a theorem of Cartan and Dieudonne, but doesn't follow from the proofs given above.
(ii) One can, somewhat analogously, prove that any self-isometry of a (V, B) with B an alternating nondegenerate bilinear form, is a product of what are called symplectic transvections: each of these, say τ, by definition fixes a hyperplane $W \subset V$, and satisfies $\tau(v) - v \in W$ for all $v \in V$. We will not get into the details.
(iii) These theorems do not work as such for quadratic forms in characteristic two, but some variants of theirs do hold: e.g., there is a form of Witt cancellation for nondegenerate quadratic forms q whose associated bilinear forms B_q are nondegenerate.

Corollary 10.29. *Consider the classical group $\mathrm{SO}_n(\mathbb{R}) = \{g \in \mathrm{GL}_n(\mathbb{R}) \mid g \cdot {}^t g = 1, \det g = 1\}$. Give it the topology induced from the embedding $\mathrm{SO}_n(\mathbb{R}) \hookrightarrow \mathrm{GL}_n(\mathbb{R}) \hookrightarrow \mathbb{R}^{n^2}$, obtained by reading the matrix entries. Then $\mathrm{SO}_n(\mathbb{R})$ is path connected.*

Proof. $\mathrm{SO}_n(R)$ is the special orthogonal group associated to the symmetric nondegenerate bilinear form on $\mathbb{R}^n$, thought of as the space of column vectors, given by the Euclidean inner product $(X, Y) \mapsto {}^t X \cdot Y$. Thus, by Corollary 10.27, any element of $\mathrm{O}_n(\mathbb{R})$ is a product of reflections (with respect to the Euclidean inner product). By determinant considerations, therefore, every element of $\mathrm{SO}_n(\mathbb{R})$ is a product of an even number of reflections (a reflection has determinant -1, while by definition, any element of $\mathrm{SO}_n(\mathbb{R})$ has determinant 1). Thus, it now suffices to show that any product of two reflections with respect to the Euclidean inner product, say $r_v \cdot r_w$, can be connected by a path to the identity. But $\operatorname{Span}(v, w)$ is equal to or contained in some two-dimensional subspace $W \subset \mathbb{R}^n$, and $r_v \cdot r_w$ then belongs to $\mathrm{SO}(W) \subset \mathrm{SO}(\mathbb{R}^n)$, where we are viewing $\mathrm{SO}(W)$ as the subgroup of $\mathrm{SO}(\mathbb{R}^n)$ acting as the identity on $W^\perp$ (use Remark 10.24). Since $\mathrm{SO}(W) \cong \mathrm{SO}(2)$ is homeomorphic to S^1 and hence path connected, $r_v r_w$ lies in the path connected component of $1 \in \mathrm{SO}_n(\mathbb{R})$. □

10.4. **Clifford algebras (Optional, not discussed in the lecture).**

Definition 10.30. Let (V, q) be a quadratic space over F. Then the Clifford algebra $C(V)$ of V is defined to be the following quotient of the tensor algebra $T(V)$ of V (over F):

$$C(V) = T(V)/I_C(V), \quad \text{where } I_C(V) = \text{the two-sided ideal generated by } \{v \otimes v - q(v) \mid v \in V\}.$$

Denote the image of $v_1 \otimes \cdots \otimes v_n \in T^n V$ in $C(V)$ by $v_1 \dots v_n$.

$I_C(V) \subset C(V)$ is not a homogeneous ideal, so $C(V)$ is not $\mathbb{Z}$-graded. However, its generators are linear combinations of elements $v \otimes v$ of degree 2 and elements $q(v)$ of degree 0, so $C(V)$ gets a $\mathbb{Z}/2\mathbb{Z}$-grading instead:

$$C(V) = C^0(V) \oplus C^1(V), \qquad \text{where } C^i(V) = \text{ the image of } \bigoplus_{n \equiv i \bmod 2} T^i(V).$$

One motivation for considering Clifford algebras is that they can be used to define a two-fold cover of special orthogonal groups, called spin groups: in appropriate settings and appropriately interpreted, these are the universal covers of special orthogonal groups. Let us assume that $\operatorname{char} F \neq 2$, so there exists a symmetric nondegenerate bilinear form B on V such that $q(v) = B(v, v)$ for all $v \in V$.

Here are some basic properties of Clifford algebras – see Professor Nair's notes or some other reference for proofs:

(i) One can show that $C(V)$ is finite dimensional over F, with dimension $2^{\dim V}$.
(ii) In fact, the "$2^{\dim V}$" can be realized as follows: if $v_1, \dots, v_n$ is an orthogonal basis for V (which exists since $\operatorname{char} F \neq 2$), then a basis for $C(V)$ is given by $\{x_{i_1} \dots x_{i_r} \mid r \leqslant n, 1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n\}$. This is not surprising: if $q = 0$, then by definition, the Clifford algebra $C(V, q)$ is just the exterior algebra $\Lambda(V)$.
(iii) The main computation in proving this is: one notes that the Clifford algebra of each Fv_i is $F[x]/(x^2 - q(v_i))$ (easy), and shows that $C((V, q) \oplus (V', q'))$ is, for a suitable notion of tensor product, $C(V, q) \otimes C(V', q')$.
(iv) From the above description for a basis of $C(V)$, it follows that the map $V = T^1(V) \to C(V)$ is an injection.

Here is some idea about why it is plausible that Clifford algebras can used to define covers of special orthogonal groups.

- If $v \in V$ is anisotropic, then in $C(V)$ we have $v \cdot v = q(v) \in F^\times \subset C^0(V)^\times$, so as an element of $C(V)$, v is invertible.
- Now let $x, v \in V$ with x anisotropic. Then in $C(V)$, $xv + vx = (x+v)^2 - x^2 - v^2 = q(x+v) - q(x) - q(v) = 2B(x, v)$. It follows that:

$$xvx^{-1} = (-vx + 2B(x,v))x^{-1} = -v + 2\frac{B(x,v)}{q(x)}x = -\Big(v - 2\frac{B(x,v)}{B(x,x)}x\Big) = -r_x(v).$$

Thus, $v \mapsto -r_x(v)$ can be realized on $V \subset C(V)$ as conjugation by x. This suggests that anisotropic vectors in $C(V)$, which we have seen to be invertible in $C(V)$, may generate something that maps to the orthogonal group.

One considers:

- Let $\varepsilon : C(V) \to C(V)$ be the automorphism that is the identity on $C^0(V)$ and acts as multiplication by -1 on $C^1(V)$: it accounts for the "$-$" in the $-r_x(v)$ above.
- One defines the Clifford group and the special Clifford group, also known as the $GPin$ group and the $GSpin$ group, to be:

$$\Gamma(V) = GPin(V) = \{x \in C(V)^\times \mid \varepsilon(x) v x^{-1} \in V \,\forall\, v \in V\}.$$

$$S\Gamma(V) = GSpin(V) = GPin(V) \cap C^0(V)^\times = \{x \in C^0(V)^\times \mid xvx^{-1} \in V \,\forall\, v \in V\}.$$

Motivation. If $x \in V \subset C(V)$ is anisotropic, then $\varepsilon(x) v x^{-1} = -xvx^{-1} = r_x(v) \in V$, so $x \in \Gamma(V)$. It acts as $r_x \in O(V, q)$ on $V \subset C(V)$.

This generalizes: one shows that sending $x \in \Gamma(V)$ to $(v \mapsto \varepsilon(x) v x^{-1}) \in GL(V)$ has image in $O(V) = O(V, q)$, [28] and gives us an exact sequence

$$1 \to F^\times \to \Gamma(V) \to O(V) \to 1$$

(these are nonabelian groups, so we haven't defined exactness: but we mean that $F^\times \to \Gamma(V)$ is injective, with image the kernel of $\Gamma(V) \to O(V)$, which is surjective). The presence of $F^\times \subset \Gamma(V)$ is not surprising: if $x \in V$ is anisotropic, then for any $\alpha \in F^\times$, x and αx are different elements in $\Gamma(V)$, but map to $r_x = r_{\alpha x}$ in $O(V)$.

This then restricts to

$$1 \to F^\times \to GSpin(V) \to SO(V) \to 1.$$

We thus get an "$F^\times$"-cover of $SO(V)$. To cut it down to a double cover, we need to remove roughly $F^\times$-worth of material from $GSpin(V)$.

- On $C(V)$, one has $x \mapsto x^*$, the unique anti-involution that is the identity on F and on V: so $(v_1 \dots v_n)^* = v_n \dots v_1$ if $v_1, \dots, v_n \in V = C(V)$. Define $N(x) = xx^*$ for $x \in C(V)$. One shows that $N : C(V) \to C(V)$ restricts to $N : \Gamma(V) \to F^\times$ (easy).
- $Spin(V)$ is defined to be $\ker(N|_{GSpin(V)}) \subset GSpin(V)$. Unsurprisingly, it maps to $SO(V)$ by $x \mapsto (V \ni v \mapsto xvx^{-1})$. Alternatively,

$$Spin(V) = \{t \in C^0(V) \mid t^* t = 1, \text{ and } tVt^{-1} = V\}.$$

Exercise 10.31. Read up about Witt rings, and more about Clifford algebras, from Professor Nair's notes.

11. A summary of very basic facts about classifying quadratic spaces

Let F be a field. We will assume that $2 \nmid \operatorname{char} F$, though it is not necessary for some of the following. We will also restrict to nondegenerate (V, q). In what follows, a lot of the results will be stated without proof. Many of the proofs can be found in notes of Professor William Casselman (Bill Casselman), whom I generally enjoy reading:

[28] This is easy if $x \in S\Gamma(V)$: $q(xvx^{-1}) = (xvx^{-1})^2 = xv^2x^{-1} = xq(v)x^{-1} = q(v)$. And not much harder if $x \in \Gamma(V) \backslash S\Gamma(V)$.

https://personal.math.ubc.ca/ cass/research/pdf/QForms.pdf

https://personal.math.ubc.ca/ cass/siegel/FiniteFields.pdf

- (V, q) with $\dim V = 1$ (and nondegenerate by the current convention): these are classified by $F^\times/F^{\times 2}$, as we saw in Lecture 9.
- (V, q) with $\dim V = 2$: A homework problem asked you to prove that these are either hyperbolic planes, or of the form $aN_{K/F}$ with $a \in F^\times$, and K/F a separable quadratic extension. Here K/F can be described as $F[\sqrt{-\det q}]$, and is hence uniquely determined up to isomorphism by q (and (V, q) is a hyperbolic plane if and only if $-\det q$ is already a square in F). It is easy to see that $aN_{K/F}$ and $a'N_{K'/F}$ are isometric if and only if K/F is isomorphic to K'/F and $a' \in a \cdot N_{K/F}(K^\times)$. This continues to be true in characteristic two.
- When F is algebraically closed: then the isometry class of (V, q) is determined by $\dim_F V$; see Corollary 10.10(ii).
- When $F = \mathbb{R}$: This is handled by Sylvester's law of inertia, see Corollary 10.10(iii). This implies that there are only finitely many $(d + 1)$ nondegenerate quadratic spaces of dimension d over $\mathbb{R}$, up to isomorphism.
- When F is a finite field: In this case, one can show that $(V, q) \cong (V', q')$ if and only if $\dim V = \dim V'$ and $\det q = \det q'$. In other words, nondegenerate quadratic forms over finite fields of odd characteristic are classified by their dimension and determinant. When $\operatorname{char} F = 2$, the determinant no longer seems to be a reasonable thing, but here is a characteristic-free description of the set of isometry classes of (V, q) with $\dim V$ equal to a given d:
  - If d is even, there are two: $\mathbb{H}^{d/2}$ and $\mathbb{H}^{(d/2)-1} \oplus N_{K/F}$, for the unique-up to-isomorphism quadratic extension K/F ($N_{K/F}$ is surjective, so considering $aN_{K/F}$ does note give us a different form).
  - If d is odd, the various $\mathbb{H}^{d/2} \oplus cx^2$, where c ranges over representatives for $F^\times/F^{\times 2}$: note that there is only one of these in characteristic two, but two in odd characteristic.
- When F is a p-adic field $\mathbb{Q}_p$: the isometry class of (V, q) is completely determined by $\dim V$, $\det q$ and what is known as a Hasse-invariant; if $q \cong \sum a_i x_i^2$, the Hasse-invariant is given by $\prod_{i<j}(a_i, a_j)$, where (x, y) stands for what is called the Hilbert symbol. Not all combinations of (dimension, discriminant, Hasse-invariant) arise, but most do. Thus, like with $\mathbb{R}$ and with finite fields, there are only finitely many isometry classes of nondegenerate quadratic forms of a given dimension.
- When $F = \mathbb{Q}$. The take home message here is that $\mathbb{Q}$ is far more complicated, for the purposes of this theory, than $\mathbb{C}, \mathbb{R}$, finite fields and p-adic fields. For instance, over $\mathbb{Q}$, even with $\dim V = 1$, there are infinitely many isomorphism classes (V, q): this is because $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ is infinite (it is the product of ± 1 and the free abelian group generated by the prime numbers). Hopefully this also gives some idea that the classification of quadratic forms over $\mathbb{Q}$ is number theoretic in nature.

One important result in the classification of quadratic forms over $\mathbb{Q}$ is the Hasse-Minkowski theorem, which says that (V, q) and (V', q') are isometric if and only if they become isomorphic when considered over $\mathbb{R}$ as well as over each $\mathbb{Q}_p$ (this helps because the theory is simpler over $\mathbb{R}$ and over each $\mathbb{Q}_p$). A proof of this result is given in Serre's "A course in arithmetic".

11. Lecture 11 — Additive and abelian categories (Incomplete/extra crude)

Today: all categories except possibly the presheaf categories that we encounter will be locally small.

Very informal motivation. The categories $AbGrp$, $R\text{-}Mod$, $Rep_k G$ etc. share many common properties: each Hom set is actually a Hom group, they all have a '0' object, the injectivity of a morphism can be checked by simply seeing if the kernel is trivial, etc.

We will define

Pre-additive categories $\rightsquigarrow$ additive categories $\rightsquigarrow$ preabelian categories $\rightsquigarrow$ abelian categories,

each more restrictive than the other.

Notation 11.1. Today, for any category $\mathcal{C}$, we will denote $\mathrm{Mor}_{\mathcal{C}}$ by $\mathrm{Hom}_{\mathcal{C}}$: this feels more natural when each $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is a group.

11.1. Additive categories.

Definition 11.2. (i) A category $\mathcal{C}$ is called preadditive if we are given the structure of an abelian group on $\mathrm{Hom}_{\mathcal{C}}(X, Y) = \mathrm{Hom}(X, Y)$ for each $X, Y \in \mathrm{Ob}\,\mathcal{C}$, such that for all $X, Y, Z \in \mathrm{Ob}\,\mathcal{C}$, the map

$$\circ : \mathrm{Hom}(Y, Z) \times \mathrm{Hom}(X, Y) \to \mathrm{Hom}(X, Z)$$

given by composition is $\mathbb{Z}$-bilinear. [29]

(ii) (a) A zero object in a category is an object which is both an initial object and a terminal object; it may not exist, but if it does it is unique up to a unique isomorphism. A (choice of a) zero object in a category $\mathcal{C}$ will be denoted by $0 = 0_{\mathcal{C}}$.

(b) If $\mathcal{C}$ has a zero object, then for all $X, Y \in \mathrm{Ob}\,\mathcal{C}$, the set $\mathrm{Hom}(X, Y)$ has a canonical element obtained as the composite $X \to 0 \to Y$. It will be called the zero morphism, and be denoted $0 : X \to Y$ or $0 \in \mathrm{Hom}(X, Y)$. Note that, in this case each $\mathrm{Hom}(X, Y)$ is a pointed set (with 0 being the distinguished element), in an appropriately functorial way, and the composite of any chain of morphisms that contains a zero object or a zero morphism, is a zero morphism.

(iii) A preadditive category $\mathcal{C}$ is called an additive category if it has a zero object and binary products. See Lemma 11.4 below for other equivalent ways to define an additive category.

We will regularly use the following from now on, mostly without further mention.

Exercise 11.3. If a preadditive category has a zero object, then the identity element of each of the groups $\mathrm{Hom}(X, Y)$ is the zero morphism.
Hint: Use the bilinearity, and the fact that each $\mathrm{Hom}(X, 0)$ and $\mathrm{Hom}(0, Y)$ are the trivial group.

[29] so not a group homomorphism or anything.

Lemma 11.4. *Let $\mathcal{C}$ be a preadditive category with a zero object. The following are equivalent (and hence imposing any of them makes $\mathcal{C}$ into an additive category):*

(i) $\mathcal{C}$ has binary products.
(ii) $\mathcal{C}$ has binary coproducts.
(iii) For each $X, Y \in \mathrm{Ob}\,\mathcal{C}$, there exist an object $X \oplus Y \in \mathrm{Ob}\,\mathcal{C}$ and morphisms

$$X \underset{\iota_X}{\overset{p_X}{\leftrightarrows}} X \oplus Y \underset{\iota_Y}{\overset{p_Y}{\leftrightarrows}} Y \ ,$$

such that

$$p_X \circ \iota_X = \mathrm{id}_X, p_Y \circ \iota_Y = \mathrm{id}_Y, p_X \circ \iota_Y = 0 \in \mathrm{Hom}(Y, X), p_Y \circ \iota_X = 0 \in \mathrm{Hom}(X, Y), \tag{41}$$

and such that

$$\iota_X \circ p_X + \iota_Y \circ p_Y = \mathrm{id}_{X \oplus Y} \, . \tag{42}$$

When these equivalent conditions hold, $(X \oplus Y, \iota_X, \iota_Y)$ is a coproduct of X and Y, while $(X \oplus Y, p_X, p_Y)$ is a product of X and Y.

Proof. If $(X \oplus Y = X \times Y, p_X, p_Y)$ is a product of X and Y, then by definition, $\exists\, \iota_X : X \to X \oplus Y$ such that $p_X \circ \iota_X = \mathrm{id}_X$, and $p_Y \circ \iota_X = 0$. Similarly define $\iota_Y : Y \to X \oplus Y$. (41) is automatically satisfied. To see (42), since $(X \oplus Y, p_X, p_Y)$ is a product of X and Y, it is enough to check that $p_X \circ (\iota_X \circ p_X + \iota_Y \circ p_Y) = p_X$, and $p_Y \circ (\iota_X \circ p_X + \iota_Y \circ p_Y) = p_Y$. The former follows since

$$p_X \circ (\iota_X \circ p_X + \iota_Y \circ p_Y) \overset{\text{bilinearity}}{=} (p_X \circ \iota_X) \circ p_X + (p_X \circ \iota_Y) \circ p_Y = \mathrm{id}_X \circ p_X + 0 \circ p_Y \overset{\text{Exercise } 11.3}{=} p_X,$$

and the latter is similar.

This gives (i) ⇒ (iii), and an appropriate variant of this argument gives (ii) ⇒ (iii).

Now suppose (iii) holds. Let us prove that $(X \oplus Y, \iota_X, \iota_Y)$ is a coproduct of X and Y, or equivalently, that for all $Z \in \mathrm{Ob}\,\mathcal{C}$:

$$\mathrm{Hom}(X \oplus Y, Z) \overset{(-\circ\iota_X, -\circ\iota_Y)}{\to} \mathrm{Hom}(X, Z) \times \mathrm{Hom}(Y, Z)$$

is a bijection. For this, let us check that a two-sided inverse is given by

$$(f, g) \mapsto f \circ p_X + g \circ p_Y.^{30}$$

Indeed, in one direction, use the computation

$$(f \circ p_X + g \circ p_Y) \circ \iota_X \overset{\text{bilinearity}}{=} f \circ (p_X \circ \iota_X) + g \circ (p_Y \circ \iota_X) = f \circ \mathrm{id}_X + g \circ 0 \overset{\text{Exercise } 11.3}{=} f \circ \mathrm{id}_X = f,$$

[30] How is this motivated? It can be helpful to think of how all this works in *AbGrp*. There, p_X and p_Y are the projections from $X \oplus Y$ onto X and Y, and the equation being quoted is saying that a morphism h on $X \oplus Y$ is the sum of $h \circ p_X$ and $h \circ p_Y$, which is obvious for *AbGrp*. Thus, one often can guess such arguments by working out the abelian group case, and then try to make the arguments 'arrow-theoretic' without involving the individual elements.

(where 'bilinearity' refers to the bilinearity of composition) and a similar computation with ι_Y. For the other direction, since for all $h \in \mathrm{Hom}(X \oplus Y, Z)$, it follows from (42) that:

$$h = h \circ \iota_X \circ p_X + h \circ \iota_Y \circ p_Y \in \mathrm{Hom}(X \oplus Y, Z).$$

This shows that ((iii)) implies (ii). A similar argument gives us that $(X \oplus Y, p_X, p_Y)$ is a product of X and Y, so (iii) implies (i).

The proof that (iii) implies the other two conditions also gives the final assertion of the theorem. □

Corollary 11.5. *In an additive category, finite coproducts and products exist coincide, i.e., if $\{X_i \mid i \in I\}$ in $\mathrm{Ob}\,\mathcal{C}$ with I finite, then $\coprod_{i\in I} X_i$ and $\prod_{i\in I} X_i$ exist, and have the same underlying object.*

Proof of Corollary 11.5. The case of the empty product is taken care of by the zero object. The case of nonempty products follows from an easy induction using Lemma 11.4. □

Definition 11.6. Given $\{X_i \mid i \in I\}$ in $\mathrm{Ob}\,\mathcal{C}$ with I finite, we will denote the objects $\coprod_{i\in I} X_i = \prod_{i\in I} X_i$ by $\bigoplus_{i\in I} X_i$, and call it the direct sum of the X_i. Since $\bigoplus_{i\in I} X_i$ is both a product and a coproduct of the X_i, it will be referred to as a biproduct of the X_i. Check (exercise!) that a biproduct of $\{X_i \mid i \in I\}$, with I finite, can also be defined as a triple $(\bigoplus_{i\in I} X_i, (\iota_i)_{i\in I}, (p_i)_{i\in I})$, where $\bigoplus_{i\in I} X_i \in \mathrm{Ob}\,\mathcal{C}$, and $\iota_j : X_j \to \bigoplus_{i\in I} X_i$ and $p_j : \bigoplus_{i\in I} X_i \to X_j$ are morphisms in $\mathcal{C}$ for each $j \in I$, subject to the conditions

$$p_i \circ \iota_j = 0 \in \mathrm{Hom}(X_j, X_i), \forall\, i \neq j, \quad p_i \circ \iota_i = \mathrm{id}_{X_i} \text{ for each } i \in I, \quad \text{and } \sum_{i\in I} \iota_i \circ p_i = \mathrm{id}_{\bigoplus_{i\in I} X_i}.$$

Example 11.7.

(i) $AbGrp, R\text{-}Mod, Mod\text{-}R, Vec_k, Vec_k^{fd}, R\text{-}Mod^{fg}$ (finitely generated R-modules), $Rep_k(G) = k[G]\text{-}Mod$ the category $TorsAbGrp$ of torsion abelian groups, the category of free R-modules, that of finitely presented R-modules etc. are all additive categories.

(ii) The category of divisible abelian groups – those abelian groups A with the property that the map $a \mapsto na$ is a surjection $A \to A$ for all $n \in \mathbb{Z}_{\geqslant 1}$ – is additive.

(iii) The categories $Ban_{\mathbb{R}}$ and $Ban_{\mathbb{C}}$ of Banach spaces over $\mathbb{R}$ and $\mathbb{C}$, and bounded linear homomorphisms between them, are additive categories. Note that in these categories, isomorphisms are not required to be norm-preserving; they just transfer the norm on the source to a norm equivalent to the one on the target. Or, one can think of Banach spaces not as complete normed linear spaces, but as topological vector spaces whose topology can be given by a complete norm.

(iv) The category $\mathbb{Z}\text{-}FilAbGrp$ of abelian groups with an increasing filtration indexed by $\mathbb{Z}$. Its objects are $(A, \{A_n\}_n)$, where A is an abelian group and $\{A_n\}_{n\in\mathbb{Z}}$ is an increasing filtration of A index by $\mathbb{Z}$, i.e., a doubly infinite sequence

$$0 \subset \cdots A_{-1} \subset A_0 \subset A_1 \subset \cdots \subset A$$

of increasing subgroups of A (as usual, define the morphisms in this category). The morphisms between $(A, \{A_n\}_n)$ and $(B, \{B_n\}_n)$ are homomorphisms $f : A \to B$

such that $f(A_n) \subset B_n$ for all n. Similarly, one can also consider filtered vector spaces.

(v) The categories Set, Top, the category of manifolds etc. are not additive categories: they don't have a zero object, since in these categories an initial object is not isomorphic to a final object.

Exercise 11.8. Supply the details justifying the claims in Example 11.7.

11.2. **Group objects.** This subsection will be slightly terse, and perhaps slightly informal. I will not ask questions in homework or examinations based on it, but I recommend your reading it, since it is a basic notion that is helpful in other topics you might learn, such as the theory of algebraic groups. Its relevance here is that it addresses the question: why not define an additive category as a category where every object has some structure that imitates the structure of abelian groups? And indeed, that is sort of possible.

Definition 11.9. A group object in a category $\mathcal{C}$ is an object $Y \in \mathrm{Ob}\,\mathcal{C}$, *together with* [31] the structure of an abelian group on $h^Y(X) = \mathrm{Hom}(X, Y)$ for each $X \in \mathrm{Ob}\,\mathcal{C}$, which is functorial in X. In other words, it consists of the object Y together with a functor $F : \mathcal{C}^{op} \rightsquigarrow Grp$, lifting $h^Y : \mathcal{C}^{op} \to Set$ in the sense that $h^Y = Forget \circ F$.

We might often refer to Y itself as the group object when the functor F is understood: this is an abuse of notation. When each $h^Y(X) = F(X)$ is an abelian group, we will informally and non-standardly refer to Y as an abelian group object.

Remark 11.10. Some people define group objects only for categories that have finite products and a final object 1. In this case, an alternate definition for a group object is as follows: it is an object $Y \in \mathrm{Ob}\,\mathcal{C}$ together with morphisms

- $m : Y \times Y \to Y$ (playing the role of group multiplication),
- $e : 1 \to Y$ (playing the role of the inclusion of the identity element in Y), and
- $inv : Y \to Y$ (playing the role of inversion in the group),

such that:

- m is associative: $m \circ (m \times \mathrm{id}_Y) = m \circ (\mathrm{id}_Y \times m) : Y \times Y \times Y \to Y$;
- e is a two-sided unit of m: $m \circ (\mathrm{id}_Y \times e) = \mathrm{pr}_1 : Y \times 1 \to Y$ and $m \circ (e \times \mathrm{id}_Y) = \mathrm{pr}_2 : 1 \times Y \to Y$;
- inv is a two-sided inverse with respect to m: $m \circ (\mathrm{id}_Y \times inv)$ and $m \circ (inv \times \mathrm{id}_Y)$ are both the composite $Y \to 1 \xrightarrow{e} Y$.

Exercise 11.11. (i) Write out in greater detail the definition for a group object in Remark 11.10, and verify the equivalence of that definition with the one in Definition 11.9.

Hint: The verification is just an application of the Yoneda lemma.

[31] The 'together with' signifies that a group object is not just an object in the category satisfying some conditions/properties, but an object *plus* some extra structure. This is like how a group is not a set with some properties, but a set *plus* some operations.

(ii) How do you express the condition of a group object being an abelian group object, using the approach of Remark 11.10?

Example 11.12. A group object in *Set* is a group, a group object in *Top* is a topological group (where we do not require topological groups to be Hausdorff; otherwise use the category of Hausdorff topological spaces), a group object in the category of manifolds is a Lie group, a group object in the category of algebraic varieties (resp., algebraic schemes) over a field k is an algebraic group (resp., algebraic group scheme) over k. Prove as many of these as you can, as an exercise.

Conclusion: Thus, given an additive category $\mathcal{C}$, every $Y \in \mathrm{Ob}\,\mathcal{C}$ can be thought of as (or rather, naturally enhanced into) an abelian group object, and the resulting map $m : Y \times Y = Y \oplus Y \to Y$ gives the group structure on $\mathrm{Hom}(X, Y)$:
(43)

$$\mathrm{Hom}(X,Y) \times \mathrm{Hom}(X,Y) \overset{\text{df. of prod.}}{=} \mathrm{Hom}(X, Y \times Y) = \mathrm{Hom}(X, Y \oplus Y) \overset{m\circ -}{\to} \mathrm{Hom}(X,Y).$$

11.3. **Additiveness is a property, not an extra structure.** We would like to describe the map $Y \oplus Y \to Y$ above; it turns out to be the codiagonal morphism defined below.

Definition 11.13. For this definition, allow $\mathcal{C}$ to be an arbitrary category.

- If $Y \in \mathrm{Ob}\,\mathcal{C}$ and $(Y \times Y, p_1, p_2)$ is a product of Y with itself, the associated diagonal morphism is the unique morphism $\Delta : Y \to Y \times Y$ defined by the requirement that $p_1 \circ \Delta = p_2 \circ \Delta = \mathrm{id}_Y$.
- Similarly, if $(Y \coprod Y, \iota_1, \iota_2)$ is a coproduct of Y with itself, the associated codiagonal morphism is the unique morphism $\nabla : Y \coprod Y \to Y$ is defined by the requirement that $\nabla \circ \iota_1 = \nabla \circ \iota_2 = \mathrm{id}_Y$.

Example 11.14. (i) In *Set* or *Top*, each diagonal $\Delta : Y \to Y \times Y$ is given by $y \mapsto (y, y)$ (hence the term 'diagonal'), while each codiagonal $\nabla : Y \coprod Y \to Y$ sends each $(y, i) \in Y \times \{0\} \cup Y \times \{1\} = Y \coprod Y$ to y.

(ii) In *AbGrp*, each diagonal $\Delta : Y \to Y \times Y$ is again given by $y \mapsto (y, y)$, but each codiagonal $\nabla : Y \coprod Y = Y \oplus Y \to Y$ is given by $(y_1, y_2) \mapsto y_1 + y_2$. This suggests that the codiagonal could give the group structures in a general additive category. This is indeed, what we are going to see below.

Lemma 11.15. *Let $\mathcal{C}$ be an additive category, and let $Y \in \mathrm{Ob}\,\mathcal{C}$. The 'multiplication' map $m : Y \oplus Y = Y \times Y \to Y$ (see around* (43) *for what this means) is the* codiagonal *morphism $Y \oplus Y \to Y$.*

For those who are not reading the subsection on group objects (Subsection 11.2) in detail, we will prove another version of Lemma 11.15 below, namely Lemma 11.16, which doesn't refer to that subsection. Lemma 11.15 is similar and simpler, though it also follows from Lemma 11.16.

Lemma 11.16. *Let $\mathcal{C}$ be an additive category, and let $X, Y \in \operatorname{Ob}\mathcal{C}$. Let (Y, p_1, p_2) be the product of Y with itself. Then the group multiplication in $\operatorname{Hom}(X,Y)$ coincides with the following composite:*
(44)

$$\operatorname{Hom}(X,Y) \times \operatorname{Hom}(X,Y) \overset{(-\circ p_1, -\circ p_2)^{-1}}{\to} \operatorname{Hom}(X, Y \times Y) = \operatorname{Hom}(X, Y \oplus Y) \overset{\nabla\circ -}{\to} \operatorname{Hom}(X,Y),$$

where $\nabla : Y \oplus Y \to Y$ is the codiagonal.

Proof. Write the coproduct of Y with itself as (Y, ι_1, ι_2).

By the $\mathbb{Z}$-bilinearity of the group structures under composition, the given composite is a group homomorphism. Therefore, to show that it coincides with the group multiplication on $\operatorname{Hom}(X,Y)$, it is enough to show that it takes each $(f,0)$ and $(0,f)$ to f.

This in turn follows if we show that, under the identification

$$\operatorname{Hom}(X,Y) \times \operatorname{Hom}(X,Y) \overset{(-\circ p_1, -\circ p_2)^{-1}}{\to} \operatorname{Hom}(X, Y \times Y) = \operatorname{Hom}(X, Y \oplus Y),$$

$(f,0)$ and $(0,f)$ correspond respectively to $\iota_1 \circ f$ and $\iota_2 \circ f$. But this is because

$$p_1 \circ \iota_1 \circ f = f, p_2 \circ \iota_1 \circ f = 0, p_1 \circ \iota_2 \circ f = 0, \quad \text{and} \quad p_2 \circ \iota_2 \circ f = f.$$

□

Exercise 11.17. Make sense of and justify the following assertion. Another way to write the description of the group multiplication on $\operatorname{Hom}(X,Y)$ in (44) is as follows: if $f, g \in \operatorname{Hom}(X,Y)$, then $f + g \in \operatorname{Hom}(X,Y)$ is given by:

$$X \overset{\Delta}{\to} X \oplus X \overset{f \oplus g}{\to} Y \oplus Y \overset{\nabla}{\to} X. \tag{45}$$

Remark 11.18. Thus, if a category $\mathcal{C}$ can be made into an additive category by putting group structures on all those $\operatorname{Hom}(X,Y)$, then there is only one way of putting those group structures, namely, given by (44), which depends only on the underlying category, and not the extra information which supposedly constituted the additive category. In other words, every additive category is uniquely determined by its underlying category.
Slogan. The additiveness of a category is a property of the category, and not extra structure.

11.4. **Additive functors.**

Definition 11.19. An additive functor is a functor $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ between additive categories $\mathcal{C}$ and $\mathcal{D}$, such that for all $X, Y \in \operatorname{Ob}\mathcal{C}$:

$$\operatorname{Hom}_{\mathcal{C}}(X,Y) \to \operatorname{Hom}_{\mathcal{D}}(F(X), F(Y)), \quad \text{given by } f \mapsto F(f),$$

is a group homomorphism.

Lemma 11.20. *If $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ is a functor, the following are equivalent:*

(i) F is additive.

(ii) *F takes a zero object in $\mathcal{C}$ to a zero object in $\mathcal{D}$, and a binary coproduct in $\mathcal{C}$ to one in $\mathcal{D}$: $F(0_{\mathcal{C}}) = 0_{\mathcal{D}}$ using obvious notation, and given a binary coproduct $(X \oplus Y, \iota_X, \iota_Y)$ of X and Y, $(F(X \oplus Y), F(\iota_X), F(\iota_Y))$ is a coproduct of $F(X)$ and $F(Y)$ (or in short, the 'obvious map' $F(X) \oplus F(Y) \to F(X \oplus Y)$ is an isomorphism);*

(iii) *F takes a zero object in $\mathcal{C}$ to a zero object in $\mathcal{D}$, and a binary product in $\mathcal{C}$ to one in $\mathcal{D}$: $F(0_{\mathcal{C}}) = 0_{\mathcal{D}}$, and given a binary product $(X \oplus Y, p_X, p_Y)$ of X and Y, $(F(X \oplus Y), F(p_X), F(p_Y))$ is a product of $F(X)$ and $F(Y)$ (or in short, the 'obvious map' $F(X \oplus Y) \to F(X) \oplus F(Y)$ is an isomorphism).*

Remark 11.21. The argument that (ii) and (iii) imply (i) was not discussed in Lecture 11; the result itself was stated at the beginning of Lecture 12. On the other hand, we will not use it in any crucial way: in fact, only to discuss some characterizations of left exactness in Lecture 12. What is given here is not a very good exposition. If you wish to follow it but find it difficult, you can ask me, or look up Lemma 12.7.1 in https://stacks.math.columbia.edu/tag/010M .

Proof of Lemma 11.20. First let us assume (i), and prove (ii) and (iii). We have

$$\mathrm{id}_{F(0_{\mathcal{C}})} = F(\mathrm{id}_{0_{\mathcal{C}}}) = F(0 \in \mathrm{Hom}_{\mathcal{C}}(0,0)) = (0 \in \mathrm{Hom}_{\mathcal{D}}(F(0), F(0)),$$

where the first step used that any functor preserves identity morphisms, while the last step used that an additive functor takes a zero morphism to a zero morphism (as follows using Exercise 11.3). Thus, the identity morphism of $F(0)$ is also a zero morphism of $F(0)$, so the claim that $F(0_{\mathcal{C}})$ is a zero object of $\mathcal{D}$ follows from the observation that any object Y whose zero morphism is also an identity morphism is a zero object: indeed, in addition to $0 \to Y \to 0$ being the identity, $Y \to 0 \to Y$ equals $0 : Y \to Y$ and hence $\mathrm{id}_Y : Y \to Y$.

Now, for the statements about coproducts and products, it is enough to prove that if $(X \oplus Y, \iota_X, \iota_Y, p_X, p_Y)$ are as in (iii) of Lemma 11.4, i.e., if these satisfy (41) and (42), then so do $(F(X \oplus Y, F(\iota_X), F(\iota_Y), F(p_X), F(p_Y))$. Since our proof so far shows that F sends a zero morphism to a zero morphism (and since it respects identity morphisms and compositions, being a functor), the condition imposed by (41) follows. The same for (42) follows from the fact that F being additive respects the '+' in it as well.

This gives both (ii) and (iii).

Now note that (ii) and (iii) are equivalent, since $F(X) \oplus F(Y) \to F(X \oplus Y) \to F(X) \oplus F(Y)$ is the identity (to see this, apply F to (41)). Therefore, it is now enough to assume both (ii) and (iii), and show (i).

We already know that F sends $0_{\mathcal{C}}$ to $0_{\mathcal{D}}$, so it remains to show that it induces a group homomorphism $\mathrm{Hom}_{\mathcal{C}}(X,Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$, i.e., that $F(f+g) = F(f) + F(g) \in \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$, for all $f, g \in \mathrm{Hom}_{\mathcal{C}}(X,Y)$. Let $(f,g) \in \mathrm{Hom}_{\mathcal{C}}(X, Y \oplus Y)$ be such that $p_1 \circ (f,g) = f$ and $p_2 \circ (f,g) = g$. Then by Lemma 11.16, $f + g = \nabla_Y \circ (f,g)$, where $\nabla_Y : Y \oplus Y \to Y$ is the codiagonal. Similarly, we have an expression $F(f) + F(g) = \nabla_{F(Y)} \circ (F(f), F(g))$.

Thus, it is enough to show that $\nabla_{F(Y)} \circ (F(f), F(g))$ equals $F(\nabla_Y \circ (f, g)) = F(\nabla_Y) \circ F(f, g)$. It follows from (iii) that the isomorphism $F(Y \oplus Y) \to F(Y) \oplus F(Y)$ transports $F(f, g)$ to $(F(f), F(g))$. Therefore, it is enough to show that the inverse isomorphism $F(Y) \oplus F(Y) \to F(Y \oplus Y)$ transports $\nabla_{F(Y)} : F(Y) \oplus F(Y) \to F(Y)$ to $F(\nabla_Y) : F(Y \oplus Y) \to F(Y)$.

The observation "$F(X) \oplus F(Y) \to F(X \oplus Y) \to F(X) \oplus F(Y)$ is the identity" gives that this inverse isomorphism is the obvious map $F(Y) \oplus F(Y) \to F(Y \oplus Y)$ (built out of $(F(\iota_1), F(\iota_2))$, and hence the claim that it transports $\nabla_{F(Y)} : F(Y) \oplus F(Y) \to F(Y)$ to $F(\nabla_Y) : F(Y \oplus Y) \to F(Y)$ follows from (ii).

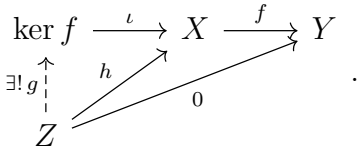
□

Example 11.22. Most of the functors we saw in the setting of R-modules, such as $\mathrm{Hom}_R(M, -), \mathrm{Hom}_R(-, M), - \otimes_R M$ etc. are all additive functors (including when R is noncommutative, in which case the target should be taken as *AbGrp*).

11.5. **Kernels and cokernels.**

Definition 11.23. Let $f : X \to Y$ be a morphism in an additive category $\mathcal{C}$.

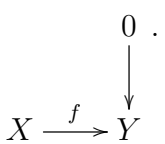
(i) A kernel of f may be defined in any of the following equivalent ways (some of these are obtained by rephrasing some others):
   (a) It is an equalizer of f and $0 : X \to Y$.
   (b) It is a pair $(\ker f, \iota)$ consisting of an object $\ker f \in \mathrm{Ob}\,\mathcal{C}$, together with a morphism $\iota : \ker f \to X$, such that for all $h : Z \to X$ in $\mathcal{C}$ with the property that $f \circ h = 0 : Z \to Y$, there exists a unique map $g : Z \to \ker(f)$ with the property that $h = \iota \circ g$:



   (c) It is an object $\ker f$ representing the functor $\mathcal{C} \to \mathit{Set}$ given by

$$Z \rightsquigarrow \ker(f \circ - : \mathrm{Hom}(Z, X) \to \mathrm{Hom}(Z, Y)),$$ [32]

   together with a natural isomorphism between $h^{\ker f}$ and this functor.
   (d) It is a limit of the diagram



[32] As usual, define it at the level of morphisms.

(ii) One similarly defines the cokernel of f, which is a coequalizer of f and 0, and is also a colimit of the diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & & \downarrow \\ & & 0 \end{array} .$$

If it exists, it corepresents the functor $\mathcal{C} \to AbGrp$ given by $Z \mapsto \operatorname{coker}(- \circ f : \operatorname{Hom}(Y, Z) \to \operatorname{Hom}(X, Z))$.

(iii) A subobject of X is a monomorphism $Z \to X$, and a quotient object of X is an epimorphism $X \to Z$. Given a subobject $Z \hookrightarrow X$, we might often write X/Z for $\operatorname{coker}(Z \hookrightarrow X)$ (in analogy with what happens for $AbGrp$).

As usual, kernels and cokernels are unique when they exist.

Exercise 11.24. Let $\mathcal{C}$ be an additive category.

(i) Given $f, g : X \to Y$ in $\mathcal{C}$, show that their equalizer is $\ker(f - g)$, and that their coequalizer if $\operatorname{coker}(f - g)$. Conclude that all equalizers exist in $\mathcal{C}$ if and only if all kernels exist in $\mathcal{C}$, and similarly with cokernels and coequalizers.

(ii) Show that $f : X \to Y$ in $\mathcal{C}$ is a monomorphism if and only if $\ker f = 0 : 0_{\mathcal{C}} \to X$, where $0_{\mathcal{C}} \in \operatorname{Ob} \mathcal{C}$ is a zero object. Prove the analogous result for epimorphisms and cokernels.

(iii) Show that every kernel is a monomorphism, and that every cokernel is an epimorphism.
Note: It may not be the case that every monomorphism is a kernel, or that every epimorphism is a cokernel. This being so is one of the characterizations of an additive category with kernels and cokernels being an abelian category, a notion we will see in the next section.

(iv) (a) In an additive category $\mathcal{C}$, if a composite $X \to Y \to Z$ is a monomorphism, show that $X \to Y$ is a monomorphism. More generally, show that $\ker(X \to Y) \to X$ factors through $\ker(X \to Y \to Z) \to X$ (and why is this "More generally"?).

(b) If this composite $X \to Y \to Z$ is an epimorphism, show that $Y \to Z$ is an epimorphism. More generally, show that $Z \to \operatorname{coker}(X \to Y \to Z)$ factors through $Z \to \operatorname{coker}(Y \to Z)$.

Note: Please do this; it is very easy.

11.6. **Preabelian categories.**

Definition 11.25. An additive category $\mathcal{C}$ is called preabelian if it satisfies:

(AB1) Kernels and cokernels exist in $\mathcal{C}$.

Example 11.26. (i) $AbGrp, R\text{-}Mod, Mod\text{-}R, Vec_k, Vec_k^{fd}$ and $Rep_k(G)$ are preabelian, where each kernel and cokernel is 'the usual ones' (or rather, the inclusion morphism of the usual kernel in the source, and the surjection to the usual cokernel from the target; in some cases we abuse notation by identifying kernels and cokernels with their underlying objects, but remember that these are really morphisms).

(ii) The full category $DivAbGrp$ of $AbGrp$ consisting of all the divisible abelian groups is preabelian: if $f : A \to B$ is a homomorphism of divisible abelian groups, the cokernel of f as computed in $AbGrp$, is divisible, and is hence also a cokernel in $DivAbGrp$. On the other hand, the kernel of f as computed in $AbGrp$ may not be divisible, but it has a 'maximal divisible subgroup', consisting of all its 'infinitely divisible' elements, which functions as a kernel in $DivAbGrp$.

(iii) The categories $Ban_{\mathbb{R}}$ and $Ban_{\mathbb{C}}$ of Banach spaces over $\mathbb{R}$ and $\mathbb{C}$, and bounded linear maps, is preabelian: given $f : X \to Y$, its kernel is the 'usual one', while its cokernel is $Y/\overline{f(X)}$, the quotient of Y by the closure of $f(X)$ in Y.

(iv) The category $\mathbb{Z}\text{-}FilAbGrp$ of filtered abelian groups is preabelian (as is, similarly, the category of filtered vector spaces over k, etc.): given $f : (A, \{A_n\}_n) \to (B, \{B_n\}_n)$, a kernel for f is $(\ker f, \{\ker f \cap A_n\}_n)$, and a cokernel for f is given by $(\operatorname{coker} f, \{\bar{B}_n\}_n)$, where $\bar{B}_n$ is the image of B_n in $\operatorname{coker} f$.

(v) The category $HTAG$ of Hausdorff topological abelian groups is preabelian: the kernels and the cokernels are as in $Ban_{\mathbb{R}}$ or $Ban_{\mathbb{C}}$.

(vi) Let R be a commutative ring (for simplicity, I guess). The category $R\text{-}Mod^{fg}$ of finitely generated R is preabelian if and only if R is Noetherian.

Preabelian categories are not enough, because often monomorphisms/epimorphisms/kernels/cokernels can behave badly:

Example 11.27. (i) In $DivAbGrp$, $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$ is both a monomorphism and an epimorphism (and thus have trivial kernel and cokernel), but it is not an isomorphism since there is no nonzero homomorphism $\mathbb{Q}/\mathbb{Z} \to \mathbb{Q}$.

(ii) In $Ban_{\mathbb{R}}$ or $Ban_{\mathbb{C}}$, if $f : X \to Y$ is injective with dense image (e.g., $C[0,1] \hookrightarrow L^2([0,1])$), then f is both a monomorphism and an epimorphism, but not an isomorphism. Similar comments apply to the category $HTAG$ of Hausdorff topological abelian groups. Alternatively, for $HTAG$, if G is a Hausdorff topological abelian group and G_d is G with discrete topology, then $G_d \to G$ is both a monomorphism and an epimorphism, but not an isomorphism.

(iii) In the category $\mathbb{Z}\text{-}FilAbGrp$ of abelian groups with an increazing $\mathbb{Z}$-filtration, consider objects $(A, \{A_n\}_n)$ and $(B, \{B_n\}_n)$, where $A = B = \mathbb{Z}$,

$$A_n = \begin{cases} 2\mathbb{Z}, & \text{if } n \geqslant 0, \\ 0, & \text{if } n < 0 \end{cases}, \qquad \text{and} \qquad B_n = \begin{cases} \mathbb{Z}, & \text{if } n \geqslant 0, \\ 0, & \text{if } n < 0 \end{cases}.$$

There is an obvious morphism $(A, \{A_n\}_n) \to (B, \{B_n\}_n)$, defined by the identity map $\mathbb{Z} \to \mathbb{Z}$, which has trivial kernel and cokernel, but is not an isomorphism.

Exercise 11.28. Justify the claims in Example 11.26 and Example 11.27.

11.7. **Abelian categories.** To define abelian categories, let us first try to define the image of a morphism $f : X \to Y$ in a preabelian category $\mathcal{C}$, using kernels and cokernels. There seem to be two obvious candidates:

Definition 11.29. Let $f : X \to Y$ be a morphism in a preabelian category $\mathcal{C}$. Then:

(i) An image of f is defined to be $\ker(\operatorname{coker} f)$: if $g : Y \to Z$ is a cokernel of f, then an image of f is, by definition, a kernel $\operatorname{im}(f) \to Y$ of g.
(ii) A coimage of f is defined to be $\operatorname{coker}(\ker f)$: if $g : Z \to X$ is a kernel of f, then a coimage of f is, by definition, a cokernel $X \to \operatorname{coim}(f)$ of g.

Note that the image and coimage of f are unique up to a unique isomorphism.

Let $K \to X$ be a kernel of f, and $Y \to C$ a cokernel of f. Write $f^\flat : X \to \operatorname{coim}(f)$ and $f^\natural : X \to \operatorname{im}(f)$ for a coimage and an image of f, respectively.

(46)
$$\begin{array}{ccccccc} K & \longrightarrow & X & \xrightarrow{\ f\ } & Y & \longrightarrow & C\ . \\ & & {\scriptstyle f^\flat}\downarrow & \nearrow_{f^\diamond} & \uparrow{\scriptstyle f^\natural} & & \\ & & \operatorname{coim}(f) & \xrightarrow{\ \bar{f}\ } & \operatorname{im}(f) & & \end{array}$$

Note that $f^\flat$, being a cokernel, is an epimorphism, while $f^\natural$, being a kernel, is a monomorphism (see Exercise 11.24(iii)).

We claim that we have a unique map $\bar{f} : \operatorname{coim}(f) \to \operatorname{im}(f)$ fitting into the above commutative diagram, i.e., such that $f = f^\natural \circ \bar{f} \circ f^\flat$. The follows because:

- Since $f \circ (K \to X) = 0$, we get $f^\diamond : \operatorname{coim}(f) \to Y$ such that $f^\diamond \circ f^\flat = f$ (as shown in the diagram).
- Since $(Y \to C) \circ f = 0$ and since $f^\flat : X \to \operatorname{coim}(f)$ is an epimorphism, , it follows that $(Y \to C) \circ f^\diamond = 0$. Since $f^\natural$ is a kernel of $Y \to C$, we get a factorization $f^\diamond = f^\natural \circ \bar{f}$. Clearly it is unique, proving the claim.

Example 11.30. Compute the following examples for images and coimages:

(i) For $AbGrp, R\text{-}Mod, Mod\text{-}R, Vec_k$ and $Rep_k(G)$, given a morphism $f : M \to N$, coim f is $M/\ker f$, while an image of f is given by the inclusion $\operatorname{im}(f) := f(M) \hookrightarrow N$, with $f(M)$ thought of as a module in the obvious way. Further, the map $\bar{f} : \operatorname{coim} f \to \operatorname{im} f$ is given by $m + \ker f \mapsto f(m)$. To say that this is an isomorphism is the first isomorphism theorem.
(ii) In $Ban_{\mathbb{R}}$ or $Ban_{\mathbb{C}}$, given $f : X \to Y$, coim $f \to$ im f identifies with the inclusion $f(X) \to \overline{f(X)}$, where $f(X)$ is made into a Banach space as the quotient of X by $\ker f$, while $\overline{f(X)}$ gets the induced Banach space structure from Y. Thus, coim $f \to$ im f is not always an isomorphism (e.g., $C[0,1] \to L^2([0,1])$). Something similar applies to the category $HTAG$ of Hausdorff topological abelian groups.

(iii) In $DivAbGrp$, given $f : X \to Y$, coim $f \to$ im f identifies with the map $X/A \to f(X)$, where A is the kernel of f in $DivAbGrp$. Thus, for $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$, coim $f \to$ im f identifies with $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$, and is hence not an isomorphism.

(iv) In $\mathbb{Z}$-$FilAbGrp$, given $f : (A, \{A_n\}_n) \to (B, \{B_n\}_n)$, im f $\to$ coim f identifies with the obvious map $(A/\ker f, \{A_n + \ker f\}_n) \to (f(A), \{f(A) \cap B_n\}_n)$. In general, the containment $f(A) \cap B_n \supset f(A_n + \ker f)$ is not an equality, so coim $f \to$ im f is not always an isomorphism.

Hopefully the above examples tell us that the following definition of an abelian category sort of amounts to imposing the first isomorphism theorem:

Definition 11.31. A category $\mathcal{C}$ is called abelian if it is preabelian, and satisfies:

(AB2) For every $f : X \to Y$ in $\mathcal{C}$, the map $\bar{f} : \mathrm{coim} f \to \mathrm{im} f$ as in (46) is an isomorphism.

Example 11.32. It follows that $AbGrp, R\text{-}Mod, Mod\text{-}R, Vec_k, Rep_k(G)$ are abelian categories, while $Ban_{\mathbb{R}}, Ban_{\mathbb{C}}, HTAG, DivAbGrp$ and $\mathbb{Z}$-$FilAbGrp$ are not abelian categories.

This can also be seen from Example 11.27; see Lemma 11.33 below.

Abelian categories avoid some of the problems in Example 11.27:

Lemma 11.33. *Any morphism $f : X \to Y$ in an abelian category that is both a monomorphism and an epimorphism, is an isomorphism.*

Proof. Since f is a monomorphism, $X \to$ coim f identifies with $\mathrm{id}_X : X \to X$, and since f is an epimorphism, im $f \to Y$ identifies with $\mathrm{id}_Y : Y \to Y$. Therefore, $\bar{f} :$ coim $f \to$ im f identifies with $f : X \to Y$, forcing f to be an isomorphism by (AB2). □

Exercise 11.34. (i) In the situation of (46), show that $\ker(\bar{f} \circ f^\flat) = \ker f$, and $\mathrm{coker}(f^\natural \circ \bar{f}) = \mathrm{coker}\, f$.

(ii) (To be checked, I haven't done this myself). In the situation of (46), show that the morphism $\bar{f}$ is both a monomorphism and an epimorphism. As we have noted before, this is not enough to ensure that $\bar{f}$ is an isomorphism.

(iii) (To be checked, I haven't done this myself). In some sources, one sees the following notions of image and coimage for a general category:

- An image of f is a monomorphism $i : I \hookrightarrow Y$ through which $f : X \to Y$ factors, and such that given any other monomorphism $i' : I' \to Y$ through which f factors, there exists a unique $v : I \to I'$ such that $i = i' \circ v$. (Think of I' as being "bigger" than I", e.g., it could be Y itself; so the image is the "smallest" monomorphism through which f factors).
- Similarly, a coimage of f is an epimorphism $c : X \to C$ through which f factors, where C is "the smallest possible", i.e., such that any other epimorphism $c' : X \to C'$ through which f factors is the composite $c' = u \circ c$ for a unique map $u : C' \to C$.

Show that for an abelian (note that I am not saying additive) category, these notions of image and coimage agree with the notions we have defined.
Note: Thus, if I understand it right, the point seems to as follows. For a general additive category, there seems to be no way to describe image and coimage using universal properties – because kernel and cokernel have sufficiently 'differently oriented' universal properties that a kernel of a cokernel or a cokernel of a kernel seems to have neither. But for an abelian category, where the coimage and the image are forced to agree, these compensate for each other's awkwardness.

Exercise 11.35. Prove, or look up a proof, that (AB2) can be described in the following equivalent ways. The following conditions on a preabelian category $\mathcal{C}$ are equivalent:

(i) It satisfies (AB2), i.e., it is an abelian category;
(ii) (Okay, this is just a rephrasing of the (AB2) given above) Given any morphism $f : X \to Y$ in $\mathcal{C}$, there exists a sequence

$$K \to X \to I \to Y \to C,$$

where:
- The composite $X \to I \to Y$ equals f;
- $K \to X$ is a kernel of f and $Y \to C$ is a cokernel of f;
- $X \to I$ is a coimage of f, and $I \to Y$ is an image of f;

(iii) Any monomorphism in $\mathcal{C}$ is a kernel, and any epimorphism in $\mathcal{C}$ is a cokernel;
(iv) Any monomorphism in $\mathcal{C}$ is the kernel of its cokernel, and any epimorphism in $\mathcal{C}$ is the cokernel of its kernel.

12. Lecture 12 — Abelian categories (contd.; incomplete/extra crude)

Today again, all categories except possibly presheaf categories will be locally small. Today we will mostly work with abelian categories, which we will denote by $\mathcal{A}, \mathcal{B}$ etc. rather than $\mathcal{C}, \mathcal{D}$ etc. Since many arguments are 'routine' but take time to write down, they will be skipped.

12.1. **Constructing abelian categories from existing ones.**

Proposition 12.1. *Let $\mathcal{B}$ be a full subcategory of an abelian category $\mathcal{A}$. Assume:*

(i) $\mathrm{Ob}\,\mathcal{B}$ *contains a zero object of* $\mathcal{A}$.
(ii) $\forall\, X, Y \in \mathrm{Ob}\,\mathcal{B}$, *some direct sum* $X \oplus Y$ *of* $X, Y \in \mathrm{Ob}\,\mathcal{A}$ *lies in* $\mathcal{B}$.
(iii) $\forall\, f : X \to Y$ *in* $\mathcal{B}$, $\mathcal{B}$ *contains some kernel of* f *in* $\mathcal{A}$, *and some cokernel of* f *in* $\mathcal{A}$.

Proof. This is easy, but let us list the steps:

- $\mathcal{B}$ is preadditive: Since $\mathcal{B}$ is a full subcategory of $\mathcal{A}$, $\mathrm{Hom}_{\mathcal{B}}(X, Y) = \mathrm{Hom}_{\mathcal{A}}(X, Y)$ is a group for all $X, Y \in \mathrm{Ob}\,\mathcal{B}$, and clearly bilinearity of composition is satisfied.
- $\mathcal{B}$ is additive: show that the zero objects and biproducts are inherited by $\mathcal{B}$ from $\mathcal{A}$.
- $\mathcal{B}$ is pre-abelian: show that for all $f : X \to Y$ in $\mathcal{B}$, any kernel or cokernel of f in $\mathcal{A}$ that lies in $\mathcal{B}$ (as is assumed to exist) also functions as a kernel or cokernel for f in $\mathcal{B}$.
- $\mathcal{B}$ is abelian: show, using the above observation on kernels and cokernels, that for all $f : X \to Y$ in $\mathcal{B}$, some coimage $X \twoheadrightarrow \mathrm{coim}(f)$ and some image $\mathrm{im}(f) \hookrightarrow Y$ of f can be taken to lie in $\mathcal{B}$, and that for these choices, the map $\bar{f} : \mathrm{coim}(f) \to \mathrm{im}(f)$ from the condition (AB2) also functions as such a map for $\mathcal{A}$, and is hence an isomorphism.

□

The following was Rishiraj's question from Lecture 11.

Proposition 12.2. *The product of two abelian categories is abelian, and in an 'obvious' way.*

Proof. Exercise. □

Proposition 12.3. *Let I be a small category and $\mathcal{A}$ an abelian category. The category* $\mathrm{Fun}(I, \mathcal{A})$ *of functors from I to $\mathcal{A}$, whose morphisms are given by natural transformations (it was introduced in Lecture 2) is an abelian category with the group laws, zero objects, biproducts, kernels and cokernels defined in an appropriately 'pointwise' sense.*

Proof. Exercise. □

Example 12.4. Here are two examples copied from Arvind's notes (for a third example of quivers, you can look up his notes):

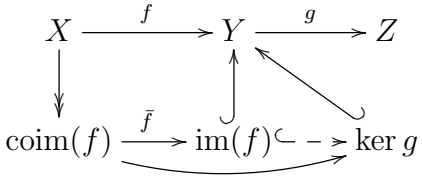
(i) Recall the category $*_G$ which has only one object, $*$, and where $\mathrm{Mor}_{\mathcal{C}}(*,*)$ is given by G. Then $Rep_k(G) = \mathrm{Fun}(*_G, Vec_k)$, recovering that $Rep_k(G)$ is an abelian category (we can also recover this by identifying $Rep_k(G)$ with $k[G]$-Mod).

(ii) If X is a topological space, let $Open(X)$ be the category of open subsets of X, where morphisms are given by inclusions. Then the category of presheaves on X with values in $\mathcal{A}$ is, by definition, the category $\mathrm{Presh}(X) := Fun(Open(X)^{op}, \mathcal{A})$, and is hence abelian. Concretely, a presheaf on X assigns to each open subset $U \subset X$ an element $\mathcal{F}_U \in \mathcal{A}$, and whenever $U \subset V \subset X$ are open subsets, we are given restriction maps $\mathcal{F}_V \to \mathcal{F}_U$, compatible with chains of inclusions $U \subset V \subset W$ (e.g., $\mathcal{A}$ could be $Vec_{\mathbb{C}}$, $\mathcal{F}_U$ could be the the vector space of continuous complex valued functions on U, and $\mathcal{F}_V \to \mathcal{F}_U$ could be the restriction of functions).

12.2. **Exactness and "**$\ker f/\mathrm{im}\ f$**".** Let $\mathcal{A}$ be an abelian category.

Definition 12.5. A chain $X \xrightarrow{f} Y \xrightarrow{g} Z$ of maps in $\mathcal{A}$ is exact if $\mathrm{im}\ f \hookrightarrow Y$ is a kernel of g.

Here is another way to put it, to help us understand it better. If $g \circ f = 0$, in *AbGrp* we have $\mathrm{im}\ f \subset \ker g$, which for abelian categories should be read as an obvious monomorphism, $\mathrm{im}\ f \hookrightarrow \ker g$:

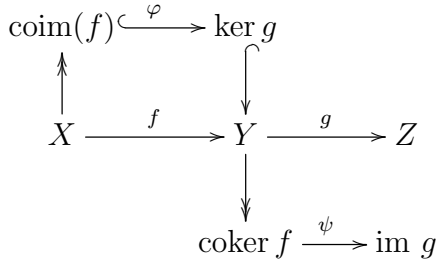


To see this, note that:

- Since $g \circ f = 0$ and since $X \to \mathrm{coim}(f)$ is an epimorphism, we have $g \circ (\mathrm{coim}(f) \to Y) = 0$ (use the definition of an epimorphism). By the universal property of $\ker g$, this naturally gives a map $\mathrm{coim}(f) \to \ker g$.
- But because $\bar{f} : \mathrm{coim}(f) \to \mathrm{im}(f)$ is an isomorphism – we are in an abelian category and not just a preabelian one – this can be promoted to a map $\mathrm{im}(f) \to \ker g$.
- This map $\mathrm{im}(f) \to \ker g$ is a monomorphism, since $\mathrm{im}(f) \to Y$ is (use one of the questions from Exercise 11.24).

Exercise 12.6. The motivation for this exercise is that in *AbGrp*, assuming $g \circ f = 0$, $(\ker g)/(\mathrm{im}\ f)$ can be described in many ways:
• $\mathrm{coker}(X \to \ker g)$; • $(\ker g)/(\mathrm{im}\ f)$; • $\mathrm{im}(\ker g \to \mathrm{coker}\ f)$; and • $\ker(\mathrm{coker}\ f \to \mathrm{im}\ g)$.
Show that these descriptions adapt to a general abelian category $\mathcal{A}$.
Hint: We will probably not need these descriptions, but perhaps this amounts to some

practice in thinking without elements. All this is described in the book of Kashiwara and Schapira, to which you can refer if the following hint is not enough. Consider diagrams:



Let u denote the composite $\ker g \hookrightarrow Y \twoheadrightarrow \operatorname{coker} f$. Show that φ is a kernel for u, and that ψ is a cokernel for u. Therefore, we get:

$$\operatorname{coker}\varphi \cong \operatorname{coim}(u) \overset{\text{(AB2)}}{\cong} \operatorname{im}(u) \cong \ker\psi. \tag{47}$$

Definition 12.7. Given a chain of maps $X \xrightarrow{f} Y \xrightarrow{g} Z$ such that $g \circ f = 0$, define

$$H(X \xrightarrow{f} Y \xrightarrow{g} Z) = \operatorname{coker}(\operatorname{im}(f) \to \ker(g)).$$

Thus, $H(X \xrightarrow{f} Y \xrightarrow{g} Z)$ also has several descriptions, as in Exercise 12.6.

We will use the following exercise, often without further comment, from now on.

Exercise 12.8. Let $\mathcal{A}$ be an abelian category. Consider chains of maps $X \xrightarrow{f} Y \xrightarrow{g} Z$ in $\mathcal{A}$ such that $g \circ f = 0$. Make these into (the objects of) a category, whose morphisms can be described by diagrams:

$$\begin{array}{ccccc} X & \xrightarrow{f} & Y & \xrightarrow{g} & Z \\ \downarrow & & \downarrow & & \downarrow \\ X' & \xrightarrow{f'} & Y' & \xrightarrow{g'} & Z' \end{array},$$

both whose squares commute. Extend $H(-)$, which is currently defined only on the objects of this category, to a functor from this category to $\mathcal{A}$.

Remark 12.9.

(i) Note that doing the above exercise involves the global axiom of choice: $H(X \xrightarrow{f} Y \xrightarrow{g} Z)$ is only well-defined up to a unique isomorphism, since $\ker g$ and $\operatorname{im}(f)$ themselves are. Thus, defining $H(-)$ as a functor involves choosing, for each object $X \xrightarrow{f} Y \xrightarrow{g} Z$ in this category, a choice of $H(X \xrightarrow{f} Y \xrightarrow{g} Z)$.

(ii) This also illustrates the importance of having things defined up to a unique isomorphism, rather than just up to an isomorphism: without things defined up to a unique isomorphism, defining $H(-)$ at the level of morphisms would involve choices, which would interfere with enuring that $H(-)$ respects compositions and the identity. Thus, things being "unique up to a unique isomorphism" is crucial to even having various functors defined.

Now, as in Lecture 1, we define:

Definition 12.10. Let $\mathcal{A}$ be an abelian category.

(i) A chain or a sequence of morphisms in A is said to be exact if it is exact at each object in the chain that is a source of some map in the chain and a target of another. Thus, $X \to Y \to Z \to W \to U$ is exact if it is exact at Y, Z and W.
(ii) A short exact sequence in $\mathcal{A}$ is a chain of morphisms in $\mathcal{A}$ of the form

$$0 \to X \to Y \to Z \to 0 \tag{48}$$

that is exact. Short exact sequences in $\mathcal{A}$ form an abelian category, which we may often denote by $SES(\mathcal{A})$.

We will use the following exercise, too, often:

Exercise 12.11. Show that (48) is exact if and only if $X \to Y$ is a monomorphism, $Y \to Z$ is an epimorphism, and the map $\mathrm{im}(f) \to \ker g$ is an isomorphism.

Remark 12.12. Any morphism $f : X \to Y$ in an abelian category 'breaks up' into two short exact sequences, as follows:

$$0 \to \ker f \to X \to \mathrm{coim}(f) = \mathrm{im}(f) \to 0,$$

and

$$0 \to \mathrm{coim}(f) = \mathrm{im}(f) \to Y \to \mathrm{coker}\, f \to 0.$$

12.3. **Left and right exactness for functors.** First, we define left and right exactness for categories that may not be abelian (but are subject to some restrictions):

Definition 12.13. (i) Let $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ be a functor, and assume that $\mathcal{C}$ has finite limits. We say that F is left exact if it preserves finite limits.
(ii) Let $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ be a functor, and assume that $\mathcal{C}$ has finite colimits. We say that F is right exact if it preserves finite colimits.

We will mainly only be interested in this definition for abelian categories (and additive functors between them), in which case we would like to make a more relatable interpretation of these notions, for which we will use the following exercise:

Exercise 12.14. (i) Assume that $\mathcal{C}$ has finite limits. Let $F : \mathcal{C} \rightsquigarrow \mathcal{D}$ be a functor. Show that the following are equivalent:
  (a) F preserves finite limits.
  (b) F preserves finite products and equalizers.
  (c) F preserves a terminal object, binary products, and equalizers.
  (d) F preserves a terminal object, and pullbacks.
(ii) Do the same for colimits. Also, do a version of all these problems without finite.

Now we give an equivalent definition for left and right exactness in the special case of abelian categories.

Definition 12.15. Let $F : \mathcal{A} \to \mathcal{B}$ be a functor.

(i) We say that F is left exact, if it is additive and preserves kernels. Note that the latter condition can be written as follows: whenever

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

is exact in $\mathcal{A}$, so is

$$0 \to F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C).$$

(ii) We say that F is right exact, if it is additive and preserves cokernels. Note that the latter condition can be written as follow: whenever

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

is exact in $\mathcal{A}$, so is

$$F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C) \to 0.$$

(iii) We say that F is exact, if it is both left exact and right exact.

Exercise 12.16. Show that Definition 12.15 is a special case of Definition 12.13.

Exercise 12.17. (Recommended)

(i) Show that a sequence

$$0 \to A \to B \to C \quad (\text{resp., } A \to B \to C \to 0)$$

in an abelian category $\mathcal{A}$ is exact if and only if for all $E \in \operatorname{Ob} \mathcal{A}$,

$$0 \to \operatorname{Hom}(E, A) \to \operatorname{Hom}(E, B) \to \operatorname{Hom}(E, C) \quad (\text{resp., } 0 \to \operatorname{Hom}(C, E) \to \operatorname{Hom}(B, E) \to \operatorname{Hom}(A, E))$$

is exact.

(ii) Let $F : \mathcal{A} \to \mathcal{B}$ be a functor between abelian categories $\mathcal{A}$ and $\mathcal{B}$, not assumed to be additive. Show that F is left-exact if and only if it preserves pull-backs, and that it is right-exact if and only if it preserves push-outs.

(iii) Let $F : \mathcal{A} \to \mathcal{B}$ be an additive functor between abelian categories $\mathcal{A}$ and $\mathcal{B}$. Show that F is exact if and only if whenever $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact in $\mathcal{A}$, $0 \to F(A) \xrightarrow{F(f)} F(B) \xrightarrow{F(g)} F(C) \to 0$ is exact in $\mathcal{B}$.

(iv) Let $\mathcal{A}$ and $\mathcal{B}$ be abelian categories, and let $F : \mathcal{A} \to \mathcal{B}$ be a functor left adjoint to a functor $G : \mathcal{B} \to \mathcal{A}$. Show that F is right exact and G is left exact (in particular, F and G are automatically additive).

12.4. **Isomorphism theorems.** We said that the condition AB2 in the definition of an abelian category was basically the first isomorphism theorem. We will now state forms of the second and the third isomorphism theorems.

In the following exercise and in what follows, whenever $Z \hookrightarrow X$ is a subobject of X, we may write X/Z for $\operatorname{coker}(Z \hookrightarrow X)$.

Proposition 12.18. *This is the second isomorphism theorem in an abelian category: Given a subobjects $M', M'' \hookrightarrow M$, define*

$$M' \cap M'' := M' \times_M M''$$[33]

(please make sure you understand that this is indeed an intersection in the context of AbGrp), and

$$M' + M'' = \mathrm{im}(M' \oplus M'' \to M).$$

Then $M'' \hookrightarrow M' \oplus M'' \to M' + M''$ is a monomorphism, and

$$M'' \hookrightarrow M' + M'' \to (M' + M'')/M'$$

factors through an isomorphism

$$M''/(M' \cap M'') \to (M' + M'')/M'.$$

Proposition 12.19. *Let*

$$0 \to M' \to M \to M'' \to 0$$

be a short exact sequence in an abelian category $\mathcal{A}$. If $L'' \hookrightarrow M''$ is a subobject, consider $L := M \times_{M''} L'' \to M$. Then $L \to L''$ is an epimorphism. Moreover, $M' \hookrightarrow M$ factors through $M' \hookrightarrow L$, and induces an isomorphism $L/M' \cong L''$.

To prove these results, we will need to understand pullbacks and pushouts in an abelian category better. This will use a few lemmas, which we will mostly not prove; but they are easy, and yet if you don't want to bother to prove them yourselves a reference for them is stacks.math.columbia.edu/tag/00ZX

Lemma 12.20. *Consider the following diagram in an abelian category $\mathcal{C}$:*

(49)
$$\begin{array}{ccc} D & \xrightarrow{k} & A \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle f} \\ B & \xrightarrow{g} & C \end{array} .$$

(i) The diagram is cartesian (i.e., realizes D as the fiber product of $f : A \to C$ and $f : B \to C$) if and only if:

$$0 \to D \overset{(k,h)}{\to} A \oplus B \overset{(f,-g)}{\to} C$$

is exact. Here, $(f, -g) : A \oplus B \to C$ is the unique map that is f when composed with $A \hookrightarrow A \oplus B$, and is $-g$ when composed with $B \hookrightarrow A \oplus B$, and (k, h) has an analogous interpretation, but with 'products' instead of 'coproducts'.

(ii) The diagram is cocartesian (i.e., realizes C as the pushout of $k : D \to A$ and $h : D \to B$) if and only if:

$$D \overset{(k,-h)}{\to} A \oplus B \overset{(f,g)}{\to} C \to 0$$

is exact.

[33] the pullback $M' \times_M M''$ exists as a special case of the fact that finite limits and colimits exist in an abelian category, something we have already seen.

A terse expression of a proof. We will prove both (i) and (ii) simultaneously. If $\mathcal{A} = AbGrp$, the lemma is easy to check, and this case is left as an exercise. We will reduce the general case to this case.

By the definition of limits, the diagram is cartesian (resp., cocartesian) if and only if for all $E \in \operatorname{Ob}\mathcal{A}$, the diagram obtained by applying $\operatorname{Hom}(E,-)$ (resp., $\operatorname{Hom}(-,E)$) to it is cartesian [34] The cartesianness of the latter diagrams, in Set, is equivalent to its cartesianness in $AbGrp$ (because fiber products work the same way for $AbGrp$ and Set). Since we know the lemma when $\mathcal{A} = AbGrp$, the cartesianness (resp., the cocartesianness) of the given diagram is therefore equivalent to the exactness, for all $E \in \operatorname{Ob}\mathcal{A}$, of

$$0 \to \operatorname{Hom}(E,D) \overset{(k\circ-,h\circ-)}{\to} \operatorname{Hom}(E,A)\oplus\operatorname{Hom}(E,B) \overset{(f,-g)\circ-)}{\to} \operatorname{Hom}(E,C)$$

$$(\text{resp., } 0 \to \operatorname{Hom}(C,E) \overset{-\circ(f,-g)}{\to} \operatorname{Hom}(A,E)\oplus\operatorname{Hom}(B,E) \overset{(-\circ k,-\circ h)}{\to} \operatorname{Hom}(D,E).)$$

By Exercise 12.17, this is equivalent to the exactness of the given sequence. □

Lemma 12.21. *(i) If* (49) *is cartesian, then the morphism* $\ker k \to \ker g$ *induced by* h *is an isomorphism.*

(ii) If (49) *is cocartesian, then the morphism* $\operatorname{coker} h \to \operatorname{coker} f$ *induced by* g *is an isomorphism.*

Remark 12.22. (i) What does h inducing $\ker k \to \ker g$ mean? For $AbGrp$ this meaning is clear, and for a general abelian category $\mathcal{A}$, we can interpret it as follows. We have $\ker k \hookrightarrow D$ and $\ker g \hookrightarrow B$, and h runs from D to B. The meaning of h inducing $\ker k \to \ker g$ is that

$$h \circ (\ker k \hookrightarrow D) : \ker k \to B$$

factors as a composite of some map $\ker k \to \ker g$ and the monomorphism $\ker g \hookrightarrow B$. The fact that $\ker g \hookrightarrow B$ is a monomorphism implies, by the definition of a monomorphism, that this factored map $\ker k \to \ker g$ is unique, allowing us to refer to it as *the* map induced by h.

(ii) Please make sure you work this lemma out in $AbGrp$ before reading the proof, and get an intuitive feel of why the statement of the lemma is reasonable.

Slightly terse proof of Lemma 12.21. If $\mathcal{A} = AbGrp$, the corollary is easy to check. We will reduce the general case to this case.

Let us prove the assertion about $\ker k \to \ker g$. For each $E \in \operatorname{Ob}\mathcal{A}$, we have a commutative diagram:

$$\begin{array}{ccccccc}
0 & \hookrightarrow & \operatorname{Hom}(E,\ker k) & \longrightarrow & \operatorname{Hom}(E,D) & \xrightarrow{k\circ-} & \operatorname{Hom}(E,A) \\
 & & \downarrow & & \downarrow{\scriptstyle h\circ-} & & \downarrow{\scriptstyle f\circ-} \\
0 & \hookrightarrow & \operatorname{Hom}(E,\ker g) & \longrightarrow & \operatorname{Hom}(E,B) & \xrightarrow{g\circ-} & \operatorname{Hom}(E,C)
\end{array},$$

[34] there is no '(resp., cocartesian)' here; limits are defined by requiring $\operatorname{Mor}(E,-)$ in Set to give a limit in Set, while colimits are defined by requiring $\operatorname{Mor}(-,E)$ to give a limit – not a colimit – in Set.

where the left vertical arrow is induced by the middle vertical one, since $\mathrm{Hom}(E, \ker k)$ identifies with $\ker(\mathrm{Hom}(E, D) \overset{k\circ -}{\to} \mathrm{Hom}(E, A))$, and similarly with $\mathrm{Hom}(E, \ker g)$. The rows of this diagram are exact, and the right square is cartesian. Since we know the case where $\mathcal{A} = AbGrp$, it follows that the left vertical arrow is an isomorphism. The Yoneda for $h_\bullet$ then gives $\ker k \to \ker g$; note that the description in the Yoneda lemma implies that $\ker k \to \ker h$ is indeed induced by h in the sense described in Remark 12.22.

The assertion about $\operatorname{coker} h \to \operatorname{coker} f$ is analogous, where one uses $\mathrm{Hom}(-, E)$ and reverses some directions instead. □

Here is a corollary to the above two lemmas:

Corollary 12.23. *(i) If (49) is cartesian and g is an epimorphism, then it is cocartesian and k is an epimorphism.*

(ii) If (49) is cocartesian and h is a monomorphism, then it is cartesian and f is a monomorphism.

Sketch of proof. If (49) is cartesian, then Lemma 12.20(i) together with the fact that g is an epimorphism implies that

$$0 \to D \overset{(k,h)}{\to} A \oplus B \overset{(f,-g)}{\to} C \to 0$$

is exact, so Lemma 12.20(ii) shows that (49) is cocartesian. That k is an epimorphism is then easy to see using Lemma 12.21.

This gives (i), and (ii) is analogous. □

Lemma 12.24. *Let $\mathcal{A}$ be an abelian category.*

(i) If $M \to N$ is an epimorphism, then for all $L \to N$, the map $M \times_N L \to L$ is an epimorphism.

(ii) If $M \to N$ is a monomorphism, then for all $M \to L$, the map $L \to L \coprod_M N$ is a monomorphism.

Proof. This is immediate from Corollary 12.23. □

Sketch of the proof of Proposition 12.18. First, $M' \hookrightarrow M$ and $M'' \hookrightarrow M$ both factor through $M' + M'' \hookrightarrow M$, and since $M' + M'' \hookrightarrow M$ is a monomorphism (being an image of a morphism to M), it is easy to see that $M' \cap M = M' \times_M M'' = M' \times_{M'+M''} M''$. [35] Moreover, since $M' \hookrightarrow M$ and $M'' \hookrightarrow M$ are monomorphisms, so are $M' \hookrightarrow M' + M''$ and $M'' \hookrightarrow M' + M''$ (use one of the questions from Exercise 11.24).

[35] In other words, if $A \to C$ and $B \to C$ factor through a monomorphism $C' \hookrightarrow C$, then we can write $A \times_C B = A \times_{C'} B$.

Thus, the following diagram is cartesian:

$$\begin{array}{ccc} M' \cap M'' & \longrightarrow & M'' \\ \downarrow & & \downarrow \\ M' & \longrightarrow & M' + M'' \end{array} .$$

If we show that this diagram is also cocartesian, it will follow from Lemma 12.21(ii) that $M'' \to M' + M''$ induces an isomorphism $M''/(M' \cap M'') \to (M' + M'')/M'$, and we would be done.

To see that this diagram is cocartesian, we combine its cartesianness with Lemma 12.20(i) to get that the following sequence is exact except possibly at $M' + M''$:

$$0 \to M' \cap M'' \to M' \oplus M'' \to M' + M'' \to 0.$$

However, $M' \oplus M'' \to M' + M''$ is the 'obvious map' $M' \oplus M'' \to M' + M''$, and hence surjective, so the above sequence is exact. Now Lemma 12.20(ii) implies that the diagram is cocartesian, as required. □

Proof of Proposition 12.19. Since $M' \to M \to M''$ equals 0, we get a unique map $M' \to L$ whose composite with $L \to M$ is the inclusion $M' \hookrightarrow M$, and whose composition with $L \to L''$ is zero. Thus, $M' \hookrightarrow M$ factors through $M' \to L$, which is a monomorphism since $M' \hookrightarrow M$ is.

By Lemma 12.21(i), applied to the cartesian diagram

$$\begin{array}{ccc} L & \longrightarrow & L'' \\ \downarrow & & \downarrow \\ M & \longrightarrow & M'' \end{array} ,$$

$L \to L''$ has $M' \to L$ as a kernel. Thus, it remains to see that $L \to L'$ is an epimorphism, which follows from Lemma 12.24. □

12.5. **Chain complexes and cochain complexes.**

Definition 12.25. Let $\mathcal{A}$ be an abelian category.

(i) A chain complex in $\mathcal{A}$ is a sequence:

$$A_\bullet : \quad \cdots \to A_{i+1} \overset{\partial_{i+1}}{\to} A_i \overset{\partial_i}{\to} A_{i-1} \overset{\partial_{i-1}}{\to} \dots$$

of morphisms in $\mathcal{A}$, indexed by $\mathbb{Z}$, such that $\partial_i \circ \partial_{i+1} = 0$ for all $i \in \mathbb{Z}$. The ∂_i are called boundary operators, or differentials. [36]

Standard abuses of notation. In what follows, if a chain complex is given as $A_\bullet, B_\bullet$ etc., we assume A_i, B_i etc. to be as above, so we have $\partial_i : A_{i+1} \to A_i, \partial_i : B_{i+1} \to B_i$

[36] Those of you who have seen homology know why the word 'boundary' is used. This is likely part of why the symbol '∂' is used, and the other probably relates to the terminology 'differentials' perhaps coming from de Rham complexes.

etc.: note that the same ∂_i be a morphism on A_i, B_i etc., so its identity will need to be figured out from the context.

(ii) A morphism $f_\bullet : A_\bullet \to B_\bullet$ of chain complexes in $\mathcal{A}$ is a collection $f_i : A_i \to B_i$ of morphisms, that are compatible with the boundary operators, i.e., such that for all $i \in \mathbb{Z}$, we have

$$\begin{array}{ccc} A_i & \xrightarrow{\partial_i} & A_{i-1} \\ \downarrow & & \downarrow \\ B_i & \xrightarrow{\partial_i} & B_{i-1} \end{array}$$

(We have made some standard abuses of notation, such as implicitly taking A_i to be as in (i) and B_i to be analogous, and writing ∂_i for the boundary operators of both $A_\bullet$ and $B_\bullet$).

(iii) With these morphisms, the chain complexes for $\mathcal{A}$ form a category, which we denote by $Ch(\mathcal{A})$ (and which we will soon see to be abelian).

(iv) A cochain complex is a sequence

$$A^\bullet : \quad \cdots \to A^{i-1} \overset{d^{i-1}}{\to} A_i \overset{d_i}{\to} A^{i+1} \overset{d^{i+1}}{\to} \dots$$

of morphisms in $\mathcal{A}$, indexed by $\mathbb{Z}$, such that $d^{i+1} \circ d^i = 0$ for all $i \in \mathbb{Z}$. [37]

(v) In obvious analogy with (ii), we define morphisms of cochain complexes in $\mathcal{A}$.

(vi) With these morphisms, cochain complexes in $\mathcal{A}$ form a category, which we will denote by $CoCh(\mathcal{A})$.

Proposition 12.26. *Let $\mathcal{A}$ be an abelian category. $Ch(\mathcal{A})$ and $CoCh(\mathcal{A})$ are abelian categories, where zero objects, biproducts, kernels and cokernels defined in an appropriately 'pointwise' sense.*

Proof. Let I be the small category with $\operatorname{Ob} I = \mathbb{Z}$, and where there exists a unique morphism $i \to j$ if $i \leqslant j$, and none otherwise. By Proposition 12.3, $\operatorname{Fun}(I^{op}, \mathcal{A})$ is abelian. This category has a description analogous to that of $Ch(\mathcal{A})$, except that the conditions $\partial_i \circ \partial_{i+1}$ are not imposed. $Ch(\mathcal{A})$ is the full subcategory of $\operatorname{Fun}(I, A)$ consisting of objects with these additional conditions imposed. Note that this subcategory, in addition to being full, is closed under zero objects, direct sums, kernels and cokernels, and is hence abelian by Proposition 12.1. □

Definition 12.27. Let $\mathcal{A}$ be an abelian category.

(i) For each $n \in \mathbb{Z}$ and each $A_\bullet \in \operatorname{Ob} Ch(\mathcal{A})$, define:

$$H_n(A_\bullet) = H(A_{n+1} \overset{\partial_{n+1}}{\to} A_n \overset{\partial_n}{\to} A_{n-1}) = \text{“}\frac{\ker \partial_n}{/}\operatorname{im}(\partial_{n+1})\text{”}$$

(see Definition 12.7). Define also the (I guess 'object of') n-cycles of $A_\bullet$, by $Z_n(A_\bullet) = \ker \partial_n \in \operatorname{Ob} \mathcal{A}$, and the n-boundaries of $A_\bullet$, by $B_n(A_\bullet) = \operatorname{im}(\partial_n)$; these are

[37] Thus, notice two difference with chain complexes: maps go from the object indexed with i to that indexed with $i+1$ rather than $i-1$, and secondly the index is superscripted rather than subscripted.

implicitly meant to be considered along with the obvious monomorphisms $Z_n \hookrightarrow A_n$ and $B_n \hookrightarrow A_n$.

(ii) For each $n \in \mathbb{Z}$ and each $A^\bullet \in \mathrm{Ob}\, CCh(\mathcal{A})$, define the degree n chain homology of $\mathcal{A}$ by

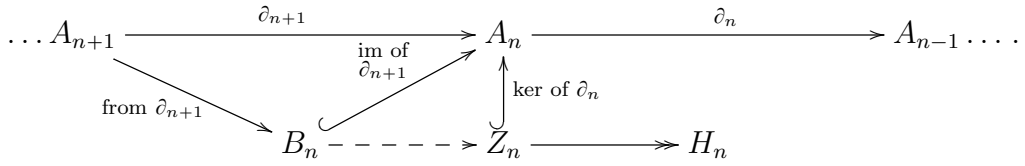
$$H^n(A^\bullet) = H(A^{n-1} \overset{d^{n-1}}{\to} A^n \overset{d^n}{\to} A^{n+1}).$$

Define also the n-cocycles of $A^\bullet$, defined by $Z^n(A^\bullet) = \ker d^n$, and the n-coboundaries of $A^\bullet$, defined by $B^n(A^\bullet) = \mathrm{im}(d^n)$. Again, these are implicitly meant to be considered with $Z^n \hookrightarrow A^n$ and $B^n \hookrightarrow A^n$.

(iii) Z_n, B_n, H_n, Z^n, B^n and H^n, defined so far at the level of objects of $Ch(\mathcal{A})$ and $Coch(\mathcal{A})$, extend to functors $Ch(\mathcal{A}) \rightsquigarrow \mathcal{A}$ and $Coch(\mathcal{A}) \rightsquigarrow \mathcal{A}$, respectively.

Exercise 12.28. (i) Justify the claim in Definition 12.27(iii).

(ii) Make sure you are clear about interpreting $H_n = Z_n/B_n$ and $H^n = Z^n/B_n$: the assertion for H_n, e.g., is that $B_n = \mathrm{im}(\partial_{n+1}) \hookrightarrow A_n$ factors through $Z_n = \ker \partial_n \hookrightarrow A_n$ (this crucially uses the condition $\partial_n \circ \partial_{n+1} = 0$), and that the resulting unique map $B_n \to Z_n$ has cokernel which is, by definition, H_n:



Example 12.29. Let R be a commutative ring.

(i) In your algebraic topology course, you will see functors

$$H_n : Top \rightsquigarrow AbGrp \qquad \text{and} \qquad H_n(-, R) : Top \rightsquigarrow R\text{-}Mod,$$

defined as composite functors

$$H_n : Top \overset{C_\bullet(-)}{\rightsquigarrow} Ch(AbGrp) \overset{H_n}{\to} AbGrp$$

(note that we are using the notation H_n stand for at least three different things, please don't get confused here), and

$$H_n(-, R) : Top \overset{C_\bullet(-,R)}{\rightsquigarrow} Ch(R\text{-}Mod) \overset{H_n}{\rightsquigarrow} R\text{-}Mod.$$

I will not describe $C_\bullet(-)$ and $C_\bullet(-, R)$, but simply say that the degree n pieces $C_n(X)$ and $C_n(X, R)$ of $C_\bullet(X)$ and $C_\bullet(X, R)$ are given as follows:

$$C_n(X) = FreeAbGrp(\{\text{Continuous maps } \Delta^n \to X\}),$$

$$C_n(X, R) = \text{the free } R\text{-module on continuous maps } \Delta^n \to X,$$

where Δ^n is the 'standard n-simplex', consisting of the convex hull of the standard basis vectors $e_0, \ldots, e_n$ in $\mathbb{R}^{n+1}$.

$C_\bullet$ is called the singular chain complex functor, $C_\bullet(X)$ the singular chain complex associated to X, $X \rightsquigarrow H_n(X)$ the singular homology functor, and $H_n(X, R)$ the singular homology of X with coefficients in the ring R.

(ii) For a topological space X, dualizing the chain complexes $C_\bullet(X)$ and $C_\bullet(X,R)$ give us *cochain complexes* $C^\bullet(X)$ and $C^\bullet(X,R)$. This gives us the singular cohomology functor

$$H^n : Top^{op} \overset{C^\bullet(-)}{\rightsquigarrow} CoCh(AbGrp) \overset{H^n}{\to} AbGrp,$$

and the functor of singular cohomology with coefficients in R:

$$H^n(-,R) : Top^{op} \overset{C^\bullet(-,R)}{\rightsquigarrow} CoCh(R\text{-}Mod) \overset{H^n}{\rightsquigarrow} R\text{-}Mod.$$

12.6. **Homological and cohomological δ-functors.** *Motivation.* Consider R-Mod, where R is a commutative ring. If $0 \to M' \to M \to M'' \to 0$ is exact, and $a \in R$ is a nonzero divisor, we know that, on taking $- \otimes_R R/(a)$, we get a right-exact sequence

$$M'/aM' \to M/aM \to M''/aM'' \to 0. \tag{50}$$

This would not be exact if we put a '$0 \to$' before M'/aM'. Therefore, we would like a means to measure a failure of exactness. For this, one uses the following exercise, which also follows from Lemma 12.32 discussed further below:

Exercise 12.30. The exact sequence (50) continues to an exact sequence

$$0 \to {}_aM' \to {}_aM \to {}_aM'' \overset{\delta}{\to} M'/aM' \to M/aM \to M''/aM'' \to 0, \tag{51}$$

where ${}_aM = \{m \in M \mid am = 0\}$, ${}_aM'$ and ${}_aM''$ are defined similarly, and the only non-obvious map $\delta : {}_aM'' \to M'/aM'$ is a map such that $\delta(m'')$ equals the image of $m' \in M'$ in M'/aM', whenever m' is chosen as follows. Since $m'' \in {}_aM''$, so that $am'' = 0$, there exists $m \in M$ with image m''. Note that $am \in \ker(M \to M'')$, so that am has some preimage in M', which we take to be m'. (In particular, you should show that given m'', though m' as above depends on the choice of m lifting m'', the image of m' in M'/aM' is independent of this choice).

Discussion on this exercise. Thus, we are not quite computing exactly the kernel of $M'/aM' \to M/aM$, which would have been ideal, but rather the best we can do in general seems to be to give an exact sequence as above; in specific situations, we would be able to manipulate such exact sequences and get whatever information we need about their kernels.

For more general left or right exact functors, we will not be able to get a six term exact sequence as above, but rather an 'infinite' exact sequence, called a long exact sequence. This is formalized in the following definition.

Definition 12.31. Let $\mathcal{A}, \mathcal{B}$ be abelian categories.

(i) A homological δ-functor from $\mathcal{A}$ to $\mathcal{B}$ is a collection of additive functors $\{T_n : \mathcal{A} \to \mathcal{B}\}_{n \geqslant 0}$, and a family of morphisms

$$\delta_n : T_n(A'') \to T_{n-1}(A')$$

for all $n \geqslant 1$, satisfying the following two conditions:

(i) For all short exact sequences $0 \to A' \xrightarrow{f} A \xrightarrow{g} A'' \to 0$, we have a long exact sequence

(52)
$$\cdots \to T_{n+1}(A'') \xrightarrow{\delta_{n+1}} T_n(A') \xrightarrow{T_n(f)} T_n(A) \xrightarrow{T_n(g)} T_n(A'') \xrightarrow{\delta_n} \ldots \xrightarrow{T_1(g)} T_1(A'') \xrightarrow{\delta_1} T_0(A') \xrightarrow{T_0(f)} T_0(A) \xrightarrow{T_0(g)} T_0(A'') \to 0.$$

(ii) Each δ_n is functorial in short exact sequences, i.e., whenever we have a map of short exact sequences:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A' & \longrightarrow & A & \longrightarrow & A'' & \longrightarrow & 0\ , \\ & & \downarrow{\scriptstyle d'} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle d''} & & \\ 0 & \longrightarrow & B' & \longrightarrow & B & \longrightarrow & B'' & \longrightarrow & 0 \end{array}$$

the following diagram commutes for each $n \geqslant 1$:

$$\begin{array}{ccc} T_n(A'') & \xrightarrow{\delta_n} & T_{n-1}(A') \\ {\scriptstyle T_n(d'')}\downarrow & & \downarrow{\scriptstyle T_{n-1}(d')} \\ T_n(B'') & \xrightarrow{\delta_n} & T_{n-1}(B') \end{array}\ .$$

In other words, this condition of 'functoriality in short exact sequences' means the following: given the functors T_n'' and T_{n-1}' on $SES(\mathcal{A})$ sending a short exact sequence $0 \to A' \to A \to A'' \to 0$ to $T_n(A'')$ and $T_{n-1}(A')$, respectively, the condition is that δ_n should define a natural transformation $T_n'' \to T_{n-1}'$.

We set $T_n = 0$ for $n < 0$.

(ii) We similarly define a cohomological δ-functor from $\mathcal{A}$ to $\mathcal{B}$ as a collection of additive functors $\{T^n : \mathcal{A} \to \mathcal{B}\}_{n \geqslant 0}$, and a family of morphisms

$$\delta^n : T^n(A'') \to T_{n+1}(A')$$

for all $n \geqslant 0$, satisfying the following two conditions:

(i) For all short exact sequences $0 \to A' \xrightarrow{f} A \xrightarrow{g} A'' \to 0$, we have a long exact sequence

(53)
$$0 \to T^0(A') \xrightarrow{T^0(f)} T^0(A) \xrightarrow{T^0(g)} T^0(A'') \xrightarrow{\delta^0} T^1(A') \xrightarrow{T^1(f)} \ldots \xrightarrow{\delta^{n-1}} T^n(A') \xrightarrow{T^n(f)} T^n(A) \xrightarrow{T^n(g)} T^n(A'') \xrightarrow{\delta^n} T^{n+1}(A') \to \ldots$$

(ii) There is an analogous condition of functoriality of δ^n in short exact sequences for each $n \geqslant 0$, taking the form of a commutative diagram:

$$\begin{array}{ccc} T^n(A'') & \xrightarrow{\delta^n} & T^{n+1}(A') \\ {\scriptstyle T^n(d'')}\downarrow & & \downarrow{\scriptstyle T^{n+1}(d')} \\ T^n(B'') & \xrightarrow{\delta^n} & T^{n+1}(B') \end{array}\ .$$

Again, we set $T^n = 0$ for $n < 0$.

Please understand the long exact sequence (52) of the above definition in analogy with Exercise 12.3051: the idea is that if a right exact functor T_0 is realized as part of a homological δ functor $((T_n)_n, (\delta_n)_n)$, then given an exact sequence $0 \to A' \to A \to A'' \to 0$, the exact sequence

$$T_0(A') \to T_0(A) \to T_0(A'') \to 0$$

is continued by (52)

Similarly, if a left exact functor T^0 is realized as a part of a cohomological δ-functor $((T^n)_n, (\delta^n)_n)$, then given an exact sequence $0 \to A' \to A \to A'' \to 0$, the exact sequence

$$0 \to T^0(A') \to T^0(A) \to T^0(A'')$$

is continued by (53).

12.7. **Snake lemma (proof mostly omitted).** The key tool in constructing long exact sequences as in (52) and (53) is the snake lemma:

Lemma 12.32. *Let $\mathcal{A}$ be an abelian category. Suppose we are given a diagram*

$$\begin{array}{ccccccc} & & M' & \xrightarrow{f} & M & \xrightarrow{g} & M'' \longrightarrow 0\ , \\ & & \downarrow{\scriptstyle d'} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle d''} \\ 0 & \longrightarrow & N' & \xrightarrow{i} & N & \xrightarrow{j} & N'' \end{array} \tag{54}$$

whose both rows are assumed to be exact. Then there is an exact sequence

$$\ker d' \xrightarrow{f} \ker d \xrightarrow{g} \ker d'' \xrightarrow{\delta} \operatorname{coker} d' \xrightarrow{\bar{i}} \operatorname{coker} d \xrightarrow{\bar{j}} \operatorname{coker} d'', \tag{55}$$

where $f, g, \bar{i}, \bar{j}$ also stand for maps induced by f, g, i, j, and δ is, informally, given by the prescription "$\delta(m'') = i^{-1} \circ d \circ g^{-1}(m'')$", which will be explicated in the proof. Moreover, this exact sequence (55) *is functorial in the diagram* (54) *(see Remark 12.34 below).*

Before proving the above lemma, let us observe that some obvious very minor variants that follow from it:

Lemma 12.33. *In* (54)*, if f is a monomorphism (resp., j is an epimorphism), then so is the map $\ker d' \xrightarrow{f} \ker d$ (resp., the map $\operatorname{coker} d \xrightarrow{\bar{g}} \operatorname{coker} d''$) of* (55)*. Thus, e.g., if f is a monomorphism and j is an epimorphism, then Lemma 12.32 implies the exactness of:*

$$0 \to \ker d' \xrightarrow{f} \ker d \xrightarrow{g} \ker d'' \xrightarrow{\delta} \operatorname{coker} d' \xrightarrow{\bar{i}} \operatorname{coker} d \xrightarrow{\bar{j}} \operatorname{coker} d'' \to 0.$$

Proof. Easy exercise. □

Some comments on the construction of δ. I will skip the proof of the snake lemma, since I don't have time. For modules it is completely straightforward, and you can do it as an easy, though tedious, exercise. For abelian categories, one needs some number of not so obvious manipulations to make the proof "element-free" and thus to work for an abelian category. Proving some of the prerequisites was the purpose of Subsection

12.4. Thus, we are not really covering this proof in the course. You can see a proof in https://stacks.math.columbia.edu/tag/00ZX or you can search for category theory notes by Julia Godecke. For a crude pointer to another approach, see Theorem 12.35 and the discussion just before it.

Nevertheless, below, I will first describe how to construct δ when $\mathcal{A} = R\text{-}Mod$, and then I will describe how to translate that into a description that works for an arbitrary abelian category.

Description of δ, when $\mathcal{A} = R\text{-}Mod$. Please compare this description to the construction of ${}_aM'' \to M'/aM'$ in Exercise 12.30.

Let $m'' \in \ker g$, and let us define $\delta(m'') \in \operatorname{coker} d'$:

- Since $M \to M''$ is surjective, $m'' = g(m)$ for some $m \in M$ (now we have "$g^{-1}(m'')$", namely m).
- Since $d''(m'') = d''g(m) = jd(m) = 0$, we have $d(m) \in \ker j$ (now we have "$dg^{-1}(m'')$", namely, $d(m)$).
- Since $\ker j = \operatorname{im}(i)$, it follows that $d(m) = i(m')$ for some $m' \in M'$ (now we have $i^{-1}dg^{-1}(m'')$, namely m').

Then $\delta(m'')$ is defined to be the image of m' in $\operatorname{coker} d' = N'/d(M')$. Of course, one needs to check that this definition is independent of the choice of m, but instead of testing the effect of replacing m, let us rewrite the above description without these choices. Namely, we have a commutative diagram

$$\begin{array}{ccccc} & & g^{-1}(\ker d'')/f(M) & = & g^{-1}(\ker d'')/(\ker g) \xrightarrow[\cong]{g} \ker d'' \ . \\ & & \downarrow{\scriptstyle d} & & \\ \operatorname{coker}(d') = N'/d'(M') & \xrightarrow[\cong]{i} & i(N')/(i\circ d')(M') & = & (\ker j)/d(f(M)) \end{array}$$

Since all maps except the vertical arrow above is an isomorphism, it gives the required map $\delta : \ker d'' \to \operatorname{coker} d'$, which agrees with the "$i^{-1}dg^{-1}(m'')$"-prescription given above.

How do we describe this in term that apply to an abelian category? For instance, $g^{-1}(\ker d'')$ can be captured as $M \times_{M''} \ker d''$. As an exercise, check that the above prescription amounts to the following: δ is the unique (as needs to be proved) dotted arrow in the following the diagram that makes it commute:

$$\begin{array}{ccccc} M & \longleftarrow & M \times_{M''} \ker d'' & \longrightarrow & \ker d'' \\ {\scriptstyle d}\downarrow & & & & \downarrow{\scriptstyle \delta} \\ N & \longrightarrow & \operatorname{coker} d' \coprod_{N'} N & \longleftarrow & \operatorname{coker} d' \end{array} \ .$$

This diagram is taken from stacks project (tag 00ZX, Section 12.5 as of typing this). You can see that page for a general proof of the snake lemma in an abelian category.

But that such a δ exists needs to be proved. I will describe how to prove that $M \times_{M''} \ker d'' \to N$ from the above diagram factors through $\ker d''$: this is about one-half the proof. Its kernel takes the form of an obvious map $M' \to M \times_{M''} \ker d''$ for a kernel, by Lemma 12.21(i). On the other hand, it is an epimorphism by Lemma 12.24(i). Thus, to show that $M \times_{M''} \ker d'' \to N$ factors through $\ker d''$, it suffices to show that

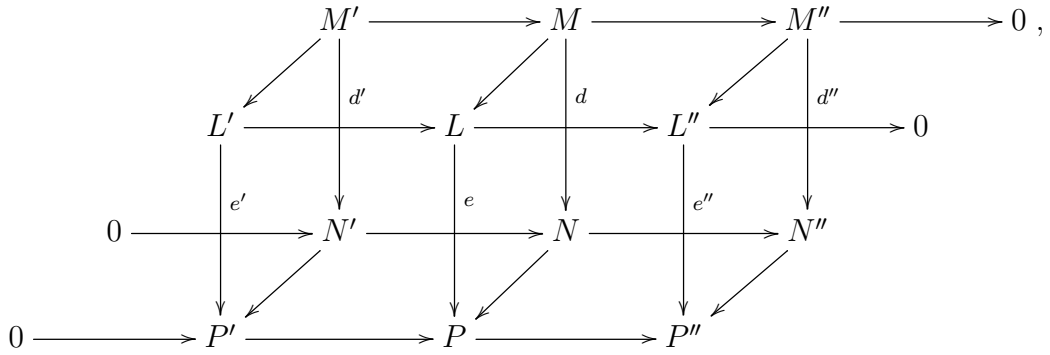
$$M' \to M \times_{M''} \ker d'' \xrightarrow{d} N \to \operatorname{coker} d' \coprod_{N'} N$$

is zero. But this chain equals, by the commutativity of the left square in the diagram:

$$(M' \xrightarrow{d'} N' \xrightarrow{i} N \to \operatorname{coker} d' \coprod_{N'} N) = (M' \xrightarrow{d'} N' \to \operatorname{coker} d' \to \operatorname{coker} d' \coprod_{N'} N) = 0,$$

where the first equality only uses the commutativity of the pushout diagram, and the second equality the fact that $M' \xrightarrow{d'} N' \to \operatorname{coker} d' = 0$. □

Remark 12.34. Now let us describe the functoriality of the snake lemma as mentioned in , e.g., (with the version of the snake lemma that does not require $f : A \to B$ to be a monomorphism) this assertion of functoriality becomes the assertion that if:



with exact rows, then there is the following commutative diagram, where both the rows are given by the snake lemma and the vertical arrows are induced by corresponding slanted arrows in the above diagram:

$$\begin{array}{ccccccccccc} \ker d' & \longrightarrow & \ker d & \longrightarrow & \ker d'' & \longrightarrow & \operatorname{coker} d' & \longrightarrow & \operatorname{coker} d & \longrightarrow & \operatorname{coker} d'' \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \ker e' & \longrightarrow & \ker e & \longrightarrow & \ker e'' & \longrightarrow & \operatorname{coker} e' & \longrightarrow & \operatorname{coker} e & \longrightarrow & \operatorname{coker} e'' \end{array} .$$

12.8. **The Freyd-Mitchell embedding theorem (statement only).** One way to prove the snake lemma for an arbitrary abelian category is to give the easy proof for R-Mod, and then appeal to the following Freyd-Mitchell embedding theorem:

Theorem 12.35 (Freyd-Mitchell embedding theorem)**.** *Let $\mathcal{A}$ be a small abelian category. Then there exists a (not necessarily commutative) ring R (with identity), and a fully faithful exact functor*

$$A \rightsquigarrow R\text{-}Mod.$$

If $\mathcal{A}$ is small, then the above theorem immediately reduces the proof of the snake lemma for $\mathcal{A}$ to that for R-Mod, which is easy to verify by hand. But even if $\mathcal{A}$ is not small, we can restrict to a suitable small abelian subcategory of $\mathcal{A}$ that contains the given diagram, and thus reduce the snake lemma for $\mathcal{A}$ to that for R-Mod.

However, this does not mean we have proved the snake lemma in general, since we have not proved the Freyd-Mitchell embedding theorem. If you are interested, a sketch of the proof is given in wikipedia.

13. Lecture 13 — Preparation for derived functors

Typically, we will denote an object of $Ch(\mathcal{A})$ by $(A_\bullet, \partial_\bullet)$, if $A_\bullet$ is given by

$$A_\bullet : \quad \ldots \overset{\partial_{i+1}}{\to} A_i \overset{\partial_i}{\to} A_{i-1} \overset{\partial_{i-1}}{\to} \ldots.$$

A morphism in $Ch(\mathcal{A})$ may be denoted by $f_\bullet : A'_\bullet \to A_\bullet$, whose the individual maps will be assumed to be denoted by $f_i : A'_i \to A_i$, as i varies over $\mathbb{Z}$. Similarly, we will write $(A^\bullet, d^\bullet)$ for an object in $Coch(\mathcal{A})$.

Throughout, given a map f in an abelian category, $\bar{f}$ will implicitly stand for a map induced by f on a subobject or a quotient object.

Remark 13.1. We will typically work with an arbitrary abelian category $\mathcal{A}$ today. Occasionally, we might explain proofs using notation that applies to R-Mod instead of to $\mathcal{A}$, but those proofs can all be adapted to the case of an arbitrary abelian category $\mathcal{A}$. If you do not feel comfortable with general abelian categories, you may, at select occasions while dealing with the proofs, just consider R-Mod instead.

13.1. The long exact sequence associated to a short exact sequence of complexes.

Remark 13.2. Since kernels and cokernels in $Ch(\mathcal{A})$ are defined "point-wise", a sequence

$$0 \to (A'_\bullet, \partial'_\bullet) \overset{f_\bullet}{\to} (A_\bullet, \partial_\bullet) \overset{g_\bullet}{\to} (A''_\bullet, \partial''_\bullet) \to 0$$

in $Ch(\mathcal{A})$ is exact if and only if for each $i \in \mathbb{Z}$,

$$0 \to A'_i \overset{f_i}{\to} A_i \overset{g_i}{\to} A''_i \to 0$$

is exact. An analogous assertion applies with $Coch(\mathcal{A})$ in place of $Ch(\mathcal{A})$.

Proposition 13.3. *Let $\mathcal{A}$ be an abelian category.*

(i) For each exact sequence

$$0 \to (A'_\bullet, \partial'_\bullet) \overset{f_\bullet}{\to} (A_\bullet, \partial_\bullet) \overset{g_\bullet}{\to} (A''_\bullet, \partial''_\bullet) \to 0 \tag{56}$$

in $Ch(\mathcal{A})$, there exist morphisms $\delta_i : H_i(A''_\bullet) \to H_{i-1}(A'_\bullet)$ for each $i \in \mathbb{Z}$, functorial in the short exact sequence (56)*, such that the following sequence is exact:*

$$\ldots \overset{\delta_{i+1}}{\to} H_i(A'_\bullet) \overset{H_i(f_\bullet)}{\to} H_i(A_\bullet) \overset{H_i(g_\bullet)}{\to} H_i(A''_\bullet) \overset{\delta_i}{\to} H_{i-1}(A'_\bullet) \overset{H_{i-1}(f_\bullet)}{\to} \ldots \tag{57}$$

(ii) For each exact sequence

$$0 \to (A^\bullet_\circ, d^\bullet_\circ) \overset{f^\bullet}{\to} (A^\bullet, d^\bullet) \overset{g^\bullet}{\to} (A^\bullet_{\circ\circ}, d^\bullet_{\circ\circ}) \to 0 \tag{58}$$

in $Coch(\mathcal{A})$, there exist morphisms $\delta^i : H^i(A^\bullet_{\circ\circ}) \to H^{i+1}(A^\bullet_\circ)$ for each $i \in \mathbb{Z}$, functorial in the short exact sequence (58)*, such that the following sequence is exact:*

$$\ldots \overset{\delta^{i-1}}{\to} H^i(A^\bullet_\circ) \overset{H^i(f_\bullet)}{\to} H^i(A^\bullet) \overset{H^i(g_\bullet)}{\to} H^i(A^\bullet_{\circ\circ}) \overset{\delta^i}{\to} H^{i+1}(A^\bullet_\circ) \overset{H^{i+1}(f_\bullet)}{\to} \ldots$$

Exercise 13.4. (i) The δ_i and the δ^i above are called connecting homomorphisms. We haven't spelled out what their functoriality in short exact sequences means. If you are uncomfortable with this, please spell it out explicitly as an exercise. It is very similar to the functoriality of the δ_n (resp., δ^n) that what we saw in the definition of a homological (resp., cohomological) δ-functor.

(ii) Consider the full (abelian) subcategory $Ch^{\geqslant 0}(\mathcal{A})$ of $Ch(\mathcal{A})$ consisting of complexes $A_\bullet$ such that $A_i = 0$ for all $i < 0$. Convince yourself that (i) of the proposition is furnishing a homological δ-functor $(\{T_n\}, \{\delta_n\})$ from $Ch^{\geqslant 0}(\mathcal{A})$ to $\mathcal{A}$, where $T_0 = H_0 : Ch^{\geqslant 0}(\mathcal{A}) \rightsquigarrow \mathcal{A}$. Similarly, (ii) is furnishing an appropriate cohomological δ-functor.

Proof of Proposition 13.3. We will prove (i); (ii) is analogous.

Applying the snake lemma to the top two rows (resp., the bottom two rows) of the diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A'_{i+1} & \xrightarrow{f_{i+1}} & A_{i+1} & \xrightarrow{g_{i+1}} & A''_{i+1} & \longrightarrow & 0\ , \\
 & & \downarrow{\scriptstyle \partial'_{i+1}} & & \downarrow{\scriptstyle \partial_{i+1}} & & \downarrow{\scriptstyle \partial''_{i+1}} & & \\
0 & \longrightarrow & A'_{i} & \xrightarrow{f_{i}} & A_{i} & \xrightarrow{g_{i}} & A''_{i} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \partial'_{i}} & & \downarrow{\scriptstyle \partial_{i}} & & \downarrow{\scriptstyle \partial''_{i}} & & \\
0 & \longrightarrow & A'_{i-1} & \xrightarrow{f_{i-1}} & A_{i-1} & \xrightarrow{g_{i-1}} & A''_{i-1} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \partial'_{i-1}} & & \downarrow{\scriptstyle \partial_{i-1}} & & \downarrow{\scriptstyle \partial''_{i-1}} & & \\
0 & \longrightarrow & A'_{i-2} & \xrightarrow{f_{i-2}} & A_{i-2} & \xrightarrow{g_{i-2}} & A''_{i-2} & \longrightarrow & 0
\end{array}$$

we get the top row (resp., the bottom row) of the following commutative diagram:

$$\begin{array}{ccccccc}
 & & \operatorname{coker} \partial'_{i+1} & \xrightarrow{\bar{f}_i} & \operatorname{coker} \partial_{i+1} & \xrightarrow{\bar{g}_i} & \operatorname{coker} \partial''_{i} \longrightarrow 0 \\
 & & \downarrow{\scriptstyle \bar{\partial}'_i} & & \downarrow{\scriptstyle \bar{\partial}_i} & & \downarrow{\scriptstyle \bar{\partial}''_i} \\
0 & \longrightarrow & \ker \partial'_{i-1} & \xrightarrow{\bar{f}_{i-1}} & \ker \partial_{i-1} & \xrightarrow{\bar{g}_{i-1}} & \ker \partial''_{i-1}
\end{array}$$

Here, the reason the vertical arrow $\bar{\partial}_i$ in the above diagram is well-defined is that $\partial_i \circ \partial_{i+1} = 0$: this ensures that ∂_i vanishes on $\operatorname{im}(\partial_{i+1}) \hookrightarrow A_i$, and hence factors as $A_i \to \operatorname{coker} \partial_{i+1} \to A_{i-1}$. [38] Similarly with $\bar{\partial}'_i$ and $\bar{\partial}''_i$.

Applying the snake lemma to this latter diagram, we get an exact sequence

$$\ker \bar{\partial}'_i \xrightarrow{\bar{f}_i} \ker \bar{\partial}_i \xrightarrow{\bar{g}_i} \ker \bar{\partial}''_i \xrightarrow{\delta_i} \operatorname{coker} \bar{\partial}'_i \xrightarrow{\bar{f}_{i-1}} \operatorname{coker} \bar{\partial}_i \xrightarrow{\bar{g}_{i-1}} \operatorname{coker} \bar{\partial}''_i \tag{59}$$

(this defines what is going to be our connecting homomorphism, δ_i).

All that remains to do now is to interpret (59). For this:

[38] This means that $\partial_i \circ (\operatorname{im}(\partial_{i+1}) \hookrightarrow A_i) = 0$, so by the universal property of the cokernel, ∂_i induces $\bar{\partial}_i : \operatorname{coker}(\partial_{i+1}) \to A_{i-1}$.

(a) Since $A_i \to \operatorname{coker} \partial_{i+1}$ is an epimorphism, it is immediate that

$$\operatorname{coker}(\bar{\partial}_i : \operatorname{coker} \partial_{i+1} \to \ker \partial_{i-1}) = \operatorname{coker}(\partial_i : A_i \to \ker \partial_{i-1}) = H_{i-1}(A_\bullet).$$

(b) Using an exercise that followed the definition of exactness in Lecture 12, we get

$$\ker(\bar{\partial}_i : \operatorname{coker} \partial_{i+1} \to \ker \partial_{i-1}) \cong H_i(A_\bullet)$$

(or, think of *R-Mod*, where this is straightforward: writing $Z_i \subset A_i$ for $\ker(\partial_i)$ and $B_i \subset A_i$ for $\operatorname{im}(\partial_{i+1})$, we are looking at

$$\ker(\bar{\partial}_i : A_i/\operatorname{im}(\partial_{i+1}) \to B_{i-1} \hookrightarrow A_{i-1}) = \ker(\bar{\partial}_i : A_i/B_i \to A_{i-1}) = Z_i/B_i = H_i(A_\bullet).$$

).

Similar assertions apply with $A_\bullet$ replaced by $A'_\bullet$ and $A''_\bullet$. With these identifications, it is easy to see that the $\bar{f}_i, \bar{g}_i, \bar{f}_{i-1}$ and $\bar{g}_{i-1}$ from (59) become, respectively, $H_i(f_\bullet), H_i(g_\bullet), H_{i-1}(f_\bullet)$ and $H_{i-1}(g_\bullet)$. Thus, (59) becomes:

$$(60) \quad H_i(A'_\bullet) \overset{H_i(f_\bullet)}{\to} H_i(A_\bullet) \overset{H_i(g_\bullet)}{\to} H_i(A''_\bullet) \overset{\delta_i}{\to} H_{i-1}(A'_\bullet) \overset{H_{i-1}(f_\bullet)}{\to} H_{i-1}(A_\bullet) \overset{H_{i-1}(g_\bullet)}{\to} H_{i-1}(A''_\bullet).$$

Now that the various δ_i are defined, so is (57). The exactness of (57) follows from the fact that (60) is exact for each i.

The functoriality of the δ_i in the short exact sequences (56) follows from the functoriality the 'δ' in the snake lemma, which is also called the connecting homomorphism. □

Remark 13.5. The above proposition doesn't tell us what the δ_n or the δ^n are, except that they can be somehow computed using the snake lemma. Yet, the above proposition is helpful, e.g.:

- Often one may know that $H_i(A''_\bullet) = 0$ for some purely formal reasons (e.g., it may be an abelian group annihilated by two coprime integers), in which case the proposition implies that $H_{i-1}(f_\bullet) : H_{i-1}(A'_\bullet) \to H_{i-1}(A_\bullet)$ is an injection. If $H_i(A''_\bullet) = H_{i-1}(A''_\bullet) = 0$, the proposition implies that this map is an isomorphism.
- The fact that the δ_n and the δ^n are functorial in short exact sequences can give a handle on them.

13.2. **A brief description of the strategy for constructing some δ-functors.** Recall that if $F : \mathcal{A} \to \mathcal{B}$ is a right-exact functor between abelian categories, we wanted to construct a homological δ-functor $(\{T_n\}, \{\delta_n\})$ with $T_0 = F$. The following is a naive strategy towards this suggested by the above proposition: it doesn't work as such, but will be modified appropriately. Try to associate to each $A \in \operatorname{Ob} \mathcal{A}$ a complex

$$A_\bullet : \ldots \overset{\partial_2}{\to} A_1 \overset{\partial_1}{\to} A_0 \to 0,$$

with the following properties:

- $A_\bullet$ should be functorial in A.

- We should have $H_0(A_\bullet) = A$, i.e., $A = \operatorname{coker}(A_1 \to A_0)$. By the right exactness of F, we would then have $F(A) = H_0(F(A_\bullet))$, where we have written $F(A_\bullet)$ somewhat abusively for the complex $\ldots \overset{F(\partial_2)}{\to} F(A_1) \overset{F(\partial_1)}{\to} F(A_0) \to 0$.
- Given a short exact sequence $0 \to A' \to A \to A'' \to 0$, the sequence $0 \to A'_\bullet \to A_\bullet \to A''_\bullet \to 0$ (resulting from the supposed functoriality of $A \rightsquigarrow A_\bullet$) should be exact in $Ch(\mathcal{A})$, i.e., exact in each degree. Moreover, so should be

$$0 \to F(A'_\bullet) \to F(A_\bullet) \to F(A''_\bullet) \to 0$$

(this is not obvious, and will depend on the A'_i, A_i, A''_i being special sorts of objects – injective or projective objects that we will see later).

In the hypothetical scenario where these conditions are satisfied, Proposition 13.3 gives a homological δ-functor $(\{H_n(F(A_\bullet)), \delta_n\})$ with $H_0(F(A_\bullet)) = F(A)$, as desired (e.g., use Exercise 13.4(i)).

It doesn't seem easy, if at all possible, to attach a single complex $A_\bullet$ to each A which satisifies all these properties. Instead, what is done is to attach to A an equivalence class of complexes $A_\bullet$ with $H_0(A_\bullet) = A$; one should then show that the choice of $A_\bullet$ in its equivalence class does not matter. The equivalence relation that is useful here turns out to be chain homotopy, which we now proceed to discuss.

13.3. **Chain and cochain homotopies.** Throughout this subsection, let $\mathcal{A}$ be an abelian category.

Definition 13.6. (i) Define a morphism $u_\bullet : (A'_\bullet, \partial'_\bullet) \to (A_\bullet, \partial_\bullet)$ in $Ch(\mathcal{A})$ to be null homotopic if there exists a sequence of morphisms $(s_i : A'_i \to A_{i+1})_{i\in\mathbb{Z}}$ such that for all $i \in \mathbb{Z}$, we have

$$u_i = \partial_{i+1} \circ s_i + s_{i-1} \circ \partial'_i.$$

$$\begin{array}{ccccccccc}
\ldots & \longrightarrow & A'_{i+1} & \xrightarrow{\partial'_{i+1}} & A'_i & \xrightarrow{\partial'_i} & A'_{i-1} & \longrightarrow & \ldots. \\
 & & \downarrow{\scriptstyle u_{i+1}} & \swarrow{\scriptstyle s_i} & \downarrow{\scriptstyle u_i} & \swarrow{\scriptstyle s_{i-1}} & \downarrow{\scriptstyle u_{i-1}} & & \\
\ldots & \longrightarrow & A_{i+1} & \xrightarrow{\partial_{i+1}} & A_i & \xrightarrow{\partial_i} & A'_{i-1} & \longrightarrow & \ldots
\end{array}$$

If $u_\bullet$ is null homotopic, we write $u_\bullet \sim 0$.

(ii) Given $u_\bullet, v_\bullet : (\mathcal{A}'_\bullet, \partial'_\bullet) \to (\mathcal{A}_\bullet, \partial_\bullet)$, we say that $u_\bullet \sim v_\bullet$, or that $u_\bullet$ and $v_\bullet$ are (chain) homotopy equivalent, if $u_\bullet - v_\bullet \sim 0$, namely, if there exists a collection $(s_i : A'_i \to A_{i+1})_{i\in\mathbb{Z}}$ of morphisms such that

$$u_i - v_i = \partial_{i+1} \circ s_i + s_{i-1} \circ \partial'_i \tag{61}$$

for all $i \in \mathbb{Z}$.

In this case, $(s_i)_i$ is said to be a (chain) homotopy equivalence between $u_\bullet$ and $v_\bullet$.

(iii) Similarly, given $u^\bullet, v^\bullet : (A^\bullet_\circ, d^\bullet_\circ) \to (A^\bullet, d^\bullet)$ in $Coch(\mathcal{A})$, we define what it means for $u^\bullet$ to be null homotopic, as denoted by $u^\bullet \sim 0$, and what it means for $u^\bullet$ to be (cochain) homotopy equivalent to $v^\bullet$, as denoted by $u^\bullet \sim v^\bullet$. Namely, the latter condition is equivalent to there existing a a collection of morphisms $(s^i : A^i_\circ \to A^{i-1})_{i\in\mathbb{Z}}$, called a (cochain) homotopy equivalence between $u^\bullet$ and $v^\bullet$, such that $u^i - v^i = d^{i-1} \circ s^i + s^{i+1} \circ d^i$ for all $i \in \mathbb{Z}$.

The definition can be motivated by the following lemma which says that often, we can pass to a homotopy class of maps.

Lemma 13.7. *(i) Chain homotopic maps induce the same maps on homology: given homotopy equivalent $u_\bullet, v_\bullet : A'_\bullet \to A_\bullet$ in $Ch(\mathcal{A})$, we have*

$$H_i(u_\bullet) = H_i(v_\bullet) : H_i(A'_\bullet) \to H_i(A_\bullet), \qquad \forall\, i \in \mathbb{Z}.$$

(ii) Cochain homotopic maps induce the same maps on cohomology: given homotopy equivalent $u^\bullet, v^\bullet : A^\bullet_\circ \to A^\bullet$ in $Coch(\mathcal{A})$, we have

$$H^i(u^\bullet) = H^i(v^\bullet) : H^i(A^\bullet_\circ) \to H^i(A^\bullet), \qquad \forall\, i \in \mathbb{Z}.$$

Proof. We prove (i); (ii) is analogous. Since $H_i(u_\bullet - v_\bullet) = H_i(u_\bullet) - H_i(v_\bullet)$ (this is entirely straightforward, but please make sure you understand this clearly), we may replace $u_\bullet$ and $v_\bullet$ by $u_\bullet - v_\bullet$ and 0, respectively. Thus, it is enough to start with some $u_\bullet \sim 0$, and show that

$$H_i(u_\bullet) : H_i(A'_\bullet) \to H_i(A_\bullet) \tag{62}$$

is the 0 morphism, for all $i \in \mathbb{Z}$. Recall the notation $Z_i = \ker(\partial_i), B_i = \mathrm{im}(\partial_i) \subset A_i$; similarly we set $Z'_i = \ker(\partial'_i), B'_i = \mathrm{im}(\partial'_i) \subset A'_i$. Recall that (62) is induced by $u_i|_{Z'_i} : Z'_i \to Z_i$, so it is enough to show that $u_i|_{Z'_i}$ factors through $B_i \hookrightarrow Z_i$. [39]

On Z'_i, we have $\partial'_i = 0$, so $u_i = \partial_{i+1} \circ s_i + s_{i-1} \circ \partial'_i$ restricts to $u_i|_{Z_i} = \partial_{i+1} \circ s_i|_{Z'_i}$, and clearly $\partial_{i+1} \circ s_i$ has image in B_i, as desired. □

Remark 13.8. Here is a motivation for the definition of homotopy equivalence, from algebraic topology, which is where it originated. A basic theorem one studies about singular homology says that homotopic maps induce the same maps on homology: if $f, g : X \to Y$ are homotopy equivalent continuous maps between topological spaces, then for all $i \in \mathbb{Z}$,

$$H_i(f) = H_i(g) : H_i(X) \to H_i(Y). \tag{63}$$

Recall that $H_i(X) = H_i(C_\bullet(X))$, where we write $C_\bullet$ for the singular chain complex functor. Thus, by Lemma 13.7, (63) follows if one shows that

$$C_\bullet(f), C_\bullet(g) : C_\bullet(X) \to C_\bullet(Y)$$

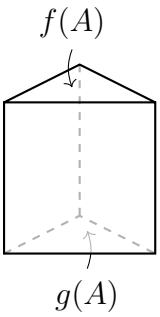
are homotopy equivalent to each other. While the proof is involved, I will make some vague remarks about it, that hopefully help motivate the above notion of chain homotopy

[39] Of course, $u_i|_{Z'_i}$ really stands for $u_i \circ (Z'_i \hookrightarrow A'_i)$.

equivalence for complexes. Consider an i-simplex A in X (namely, a continuous map $\Delta^i \to X$), and its images $f(A)$ and $g(A)$ under f and g in Y, which are simplices in Y. The homotopy itself is a map $[0,1] \times X \to Y$, so what that gives on applying to the simplex A is not a simplex in Y, but a *prism* in Y, with the simplices $f(A)$ and $g(A)$ as faces (see the figure below). The point is that while the difference between $f(A)$ and $g(A)$ is not the boundary of the prism, (as you can see in the figure below) their difference can be described (up to putting in appropriate signs) as:

$$\text{(the boundary of the prism)} - \text{(the prism over the boundary)}$$

– namely, the top and the bottom faces of the prism in the following figure are obtained by subtracting, from the collection of all faces (the boundary of the prism), just the "side faces" (the prism over the boundary, which is a prism over a triangle in the picture).



This is what is reflected in (61): s_i stands for 'taking the prism on a given simplex' (which is why it goes from A'_i to A_{i+1}, one dimension higher), $\partial_{i+1} \circ s_i$ stands for the 'boundary of the prism' ('∂' is boundary), and $s_{i-1} \circ \partial_i$ stands for the 'prism over the boundary'.

Definition 13.9. The above definition motivates associating to each additive category $\mathcal{A}$ the following categories.

(i) The homotopy category of chain complexes in $\mathcal{A}$, denoted by $Kch(\mathcal{A})$, is defined as follows:
  (a) $\operatorname{Ob} Kch(\mathcal{A}) = \operatorname{Ob} Ch(\mathcal{A})$.
  (b) $\operatorname{Hom}_{Kch(\mathcal{A})}(A'_\bullet, A_\bullet) = \operatorname{Hom}_{Ch(\mathcal{A})}(A'_\bullet, A_\bullet)/\sim$, where $\sim$ stands for homotopy equivalence.

  To really complete this definition, we must have a well-defined composition of morphisms, which is described in Exercise 13.10 below.

(ii) The homotopy category of cochain complexes in $\mathcal{A}$, denoted by $Kcoch(\mathcal{A})$, is defined as follows:
  (a) $\operatorname{Ob} Kcoch(\mathcal{A}) = \operatorname{Ob} Coch(\mathcal{A})$.
  (b) $\operatorname{Hom}_{Kcoch(\mathcal{A})}(A^\bullet_\circ, A^\bullet) = \operatorname{Hom}_{Coch(\mathcal{A})}(A^\bullet_\circ, A^\bullet)/\sim$, where $\sim$ stands for homotopy equivalence.

  Again, the composition of morphisms is left to Exercise 13.10 below.

Exercise 13.10. (i) Show that if $u_\bullet, v_\bullet : A'_\bullet \to A_\bullet$ are chain homotopic, as are $w_\bullet, t_\bullet : A_\bullet \to A''_\bullet$, then so are $w_\bullet \circ u_\bullet$ and $t_\bullet \circ v_\bullet$. Conclude that the composition in

$Ch(\mathcal{A})$ induces a well-defined composition in $Kch(\mathcal{A})$ (once we see that it is well-defined, the associativity of multiplication and the existence of identity morphisms are immediate).

(ii) Similarly, show that the composition in $Coch(\mathcal{A})$ induces a well-defined composition in $Kcoch(\mathcal{A})$.

(iii) Show that $Kch(\mathcal{A})$ and $Kcoch(\mathcal{A})$ are additive categories.
Note: It is a fact that they are not abelian categories, but we will not even discuss an example to illustrate this, for now. But if we discuss triangulated categories later, we may do so then.

We have an obvious functor $Ch(\mathcal{A}) \to Kch(\mathcal{A})$, taking each $A_\bullet \in \mathrm{Ob}\, Ch(\mathcal{A})$ to itself, and each morphism $u_\bullet$ in $Ch(\mathcal{A})$ to its chain homotopy class. Similarly, we have an obvious functor $Coch(\mathcal{A}) \to Kcoch(\mathcal{A})$. Now Lemma 13.7 implies:

Proposition 13.11. *For all $i \in \mathbb{Z}$, $H_i : Ch(\mathcal{A}) \to \mathcal{A}$ factors as a composite*

$$Ch(\mathcal{A}) \to Kch(\mathcal{A}) \stackrel{H_i}{\to} \mathcal{A}$$

(note that we are using H_i to also denoted the factored functor $Kch(\mathcal{A}) \to \mathcal{A}$). Similarly, $H^i : Coch(\mathcal{A}) \to \mathcal{A}$ factors as a composite

$$Coch(\mathcal{A}) \to Kcoch(\mathcal{A}) \stackrel{H^i}{\to} \mathcal{A}.$$

Proof. This is just a restatement of Lemma 13.7. □

Now, to carry out (the appropriate modification of) the strategy in Subsection 13.2, we will construct – under some assumptions – functors

$$\mathcal{A} \to Kch(\mathcal{A}) \stackrel{H_i}{\to} \mathcal{A}, \quad \mathcal{A} \to Kcoch(\mathcal{A}) \stackrel{H^i}{\to} \mathcal{A}.$$

Constructing the functors $\mathcal{A} \to Kch(\mathcal{A})$ and $\mathcal{A} \to Kcoch(\mathcal{A})$ will need the notion of injective and projective objects.

13.4. **Projective and injective objects.** In this subsection, let $\mathcal{A}$ denote an arbitrary abelian category.

Definition 13.12. (i) An object P in the abelian category $\mathcal{A}$ is called projective if $\mathrm{Hom}(P, -) : \mathcal{A} \rightsquigarrow AbGrp$ is exact.

(ii) An object I in $\mathcal{A}$ is called injective if $\mathrm{Hom}(-, I) : \mathcal{A}^{op} \rightsquigarrow AbGrp$ is exact.

Exercise 13.13. It was an exercise from Lecture 12 that if $\mathcal{A}$ is abelian, then so is $\mathcal{A}^{op}$, in a compatible way: this is why the above definition of an injective object makes sense (i.e., we can talk of exactness for a functor on $\mathcal{A}^{op}$). Show that an object P is projective in $\mathcal{A}$ if and only if, when viewed as an element of $\mathcal{A}^{op}$, it is injective.

Proposition 13.14. *(i) For $P \in \mathrm{Ob}\, \mathcal{A}$, the following are equivalent:*

(a) P is a projective object.

(b) *If $A \to B$ is an epimorphism in A, then* $\mathrm{Hom}(P, A) \to \mathrm{Hom}(P, B)$ *is surjective: in other words, given an epimorphism $A \to B$ in $\mathcal{A}$, any morphism $P \to B$ can be lifted to a morphism $P \to A$:*

$$\begin{array}{ccccc} A & \longrightarrow & B & \longrightarrow & 0\ . \\ & \nwarrow^{\exists} & \uparrow & & \\ & & P & & \end{array}$$

(c) *Any epimorphism $p : A \to P$ has a section, i.e., there exists a morphism $s : P \to A$ such that $p \circ s = \mathrm{id}_P$.*

(ii) *For $I \in \mathrm{Ob}\,\mathcal{A}$, the following are equivalent:*

(a) *I is an injective object.*

(b) *If $A \to B$ is a monomorphism in A, then* $\mathrm{Hom}(B, I) \to \mathrm{Hom}(A, I)$ *(the restriction, or rather composition with $A \to B$) is surjective: in other words, given a monomorphism $A \to B$ in $\mathcal{A}$, any morphism $A \to I$ can be extended to a morphism $B \to I$:*

$$\begin{array}{ccccc} 0 & \longrightarrow & A & \longrightarrow & B\ . \\ & & \downarrow & \swarrow^{\exists} & \\ & & I & & \end{array}$$

(c) *Any monomorphism $\iota : I \to A$ has a section, i.e., there exists a morphism $s : A \to I$ such that $s \circ \iota = \mathrm{id}_I$.*

Before proving the proposition, let us make sure we understand the conditions (i)(c) and (ii)(c); for the following exercise, P need not be projective, and I need not be injective.

Exercise 13.15. (i) An epimorphism $p : A \to P$ in $\mathcal{A}$ is said to be a split epimorphism if it has a section, i.e., a map $s : P \to A$ such that $p \circ s = \mathrm{id}_P$. Show that p is a split epimorphism if and only if there exists an isomorphism $a : Q \oplus P \xrightarrow{\cong} A$ such that $Q \oplus P \xrightarrow{a} A \xrightarrow{p} P$ is the identity on P and 0 on Q (thus, $Q \cong \ker p$). In other words (slightly informally), a split epimorphism $p : A \to P$ is one that lets us realize the quotient P of A as a direct summand of A.
Hint: For "⇒", let $Q = \ker p$, let a be $(\mathrm{incl.}, s)$, and show that an inverse to a is given by $(\mathrm{id}_A - s \circ p, p)$.

(ii) Similarly, a monomorphism $\iota : I \hookrightarrow A$ is said to be a split monomorphism if it has a section $s : A \to I$. Show that ι is a split monomorphism if and only if there exists an isomorphism $A \cong Q \oplus I$, such that $I \hookrightarrow A \cong Q \oplus I$ is just the inclusion onto the second factor. Thus, a monomorphism $I \hookrightarrow A$ is a split monomorphism if and only if it realizes the subobject I of A as a direct summand of A.

Proof of Proposition 13.14. We will prove (i); (ii) is similar.

$\mathrm{Hom}(P, -)$ is anyway left exact, so $\mathrm{Hom}(P, -)$ is exact if and only if it satisfies the additional property of preserving epimorphisms. This gives the equivalence of (a) and (b).

If (b) is satisfied, then applying it with $B = P$ and considering the identity morphism $P \to B = P$, we get (c). It is now enough to assume (c) and prove (b). Thus, assume given an epimorphism $A \twoheadrightarrow B$, and a morphism $P \to B$. Tautologically, this morphism can be lifted to $P \to A$ if and only if $P \times_B A \to P$ has a section: this is simply the definition of the fiber product $P \times_B A$.

Thus, by the condition (c) being assumed, it is now enough to show that $P \times_B A \to P = P \times_B B$ is an epimorphism. This is a consequence of a lemma discussed in the notes to Lecture 12: [40] basically, the lemma characterizing pull-backs in an abelian category gave us the exactness of $0 \to P \times_B A \to P \oplus A \to B$, and the map $P \oplus A \to B$ is an epimorphism because $A \to B$ is. Now the lemma characterizing push-outs in an abelian category showed us that this realizes B as a push-out of P and A over $P \times_B A$; now use the lemma which said that pushout preserves cokernels). □

Exercise 13.16. If you are not comfortable with the implication proof of the implication (c) ⇒ (b) of Proposition 13.14, prove this implication when $\mathcal{A} = R\text{-}Mod$.

Lemma 13.17. *If $P = \bigoplus_{i \in I} P_i$ in $\mathcal{A}$, then P is projective if and only if each P_i is projective. If $I = \prod_{i \in I} I_i$ in $\mathcal{A}$, then I is injective if and only if each I_i is injective.*

Proof. For the assertion about projective objects, use the criterion in (i)(b) of Proposition 13.14, as follows. Given an epimorphism $A \to B$ in $\mathcal{A}$, $\mathrm{Hom}(P, A) \to \mathrm{Hom}(P, B)$ identifies with

$$\prod_i \mathrm{Hom}(P_i, A) \cong \mathrm{Hom}\Big(\bigoplus_i P_i, A\Big) = \mathrm{Hom}(P, A) \to \mathrm{Hom}(P, B) = \mathrm{Hom}\Big(\bigoplus_i P_i, B\Big) \cong \prod_i \mathrm{Hom}(P_i, B)$$

(use that Hom from a direct sum (colimit) is the direct product of the Hom's (limit)). Thus, $\mathrm{Hom}(P, A) \to \mathrm{Hom}(P, B)$ is surjective if and only if each $\mathrm{Hom}(P_i, A) \to \mathrm{Hom}(P_i, B)$ is. The proof of the assertion about injective objects is similar, using the criterion (ii)(b) of Proposition 13.14. □

The conditions under which we will define homological or cohomological δ-functors are as follows.

Definition 13.18.
(i) We say that $\mathcal{A}$ has enough projectives if for each $A \in \mathrm{Ob}\,\mathcal{A}$, there exists an epimorphism $P \twoheadrightarrow A$ in $\mathcal{A}$ with $P \in \mathrm{Ob}\,\mathcal{A}$ projective.
(ii) We say that $\mathcal{A}$ has enough injectives if for each $A \in \mathrm{Ob}\,\mathcal{A}$, there exists a monomorphism $A \hookrightarrow I$ in $\mathcal{A}$ with $I \in \mathrm{Ob}\,\mathcal{A}$ injective.

13.5. **Projective and injective modules.** By a projective (resp., injective) left or right R-module, we mean a projective or an injective object in $R\text{-}Mod$ or $Mod\text{-}R$. In this subsection, let R be a ring (not necessarily commutative, with identity). For brevity, in this subsection, we may, unless otherwise stated, write 'R-module' or even just 'module'

[40] I am not sure you can trust that part of those notes of mine, but this is tag 05PK/Lemma 12.5.14 of Homological Algebra in the stacks project.

to mean 'left R-module'. Further, it will be clear, and understood to be understood, that analogous results apply to right R-modules.

Proposition 13.19. *Let P be a left R-module. Then P is projective if and only if P is a direct summand of a free left R-module, i.e., $F = P \oplus P'$ for some free left R-module F and some left R-module P'.*

Proof. "$\Rightarrow$": If P is projective, choose any surjection $F \to P$, where F is a free left R-module. That this epimorphism splits implies that $F \cong P \oplus P'$, where $P' = \ker(F \to P)$ (use Exercise 13.15).

"$\Leftarrow$": Assume that P is a direct summand of a free left R-module. Let us show that P is projective.
Case 1. $P = R$. Then P is projective, since $\mathrm{Hom}_R(R, -)$ is the forgetful functor R-$Mod \rightsquigarrow AbGrp$, and is hence exact.
Case 2. P is free. Case 1 and Lemma 13.17 together implies this case.
Case 3. General case. Let $F \cong P \oplus P'$, with F a free left R-module. F is projective by Case 2, so Lemma 13.17 implies that P is projective. □

Corollary 13.20. *R-Mod has enough projectives.*

Proof. Given M, take any surjection [41] $F \to M$ with F a free left R-module. F is projective by Proposition 13.19, and we are done. □

Proposition 13.21. *Any projective left R-module is flat, i.e., if P is a projective left R-module, then $- \otimes_R P : Mod\text{-}R \rightsquigarrow AbGrp$ is exact.*

Proof. By the right exactness of $- \otimes_R P$, it is enough to show that it preserves monomorphisms. Since tensor products commute with direct sums, it is easy to see that $\bigoplus_i M_i$ is flat if and only if each M_i is flat (i.e., the flat analogue of Lemma 13.17 holds). By Proposition 13.19, we are therefore reduced to showing that free left R-modules are flat. This might have been a homework problem, but in any case: another application of the just-mentioned flat analogue of Lemma 13.17 reduces this to showing that R is a flat left R-module, which follows from the fact that $- \otimes_R R : Mod\text{-}R \rightsquigarrow AbGrp$ is just the forgetful functor, which is exact. □

Exercise 13.22. Thus, we have already seen that

$$\text{Free} \Rightarrow \text{projective} \Rightarrow \text{flat} \Rightarrow \text{torsion free},$$

where to make sense of the last implication we assume that R is commutative. Show that the converse doesn't hold for any of these implications.
Hint: Among $\mathbb{Z}/6\mathbb{Z}$-modules, we have a direct sum decomposition $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$, so the $\mathbb{Z}/6\mathbb{Z}$-module $\mathbb{Z}/2\mathbb{Z}$ is projective, but clearly not free. $\mathbb{Q}$ is a flat $\mathbb{Z}$-module but not projective. An example of a non-flat torsion-free module was a recommended problem in an earlier homework.

[41] Of course, we mean surjective homomorphism

Exercise 13.23. Recall that for finitely generated modules over a PID, torsion-free was equivalent to free. Using this or otherwise, show that a module over a PID is flat if and only if it is torsion-free.
Hint: Write a general module M as a direct limit of its finitely generated submodules M_i, and show that if each M_i is flat then M is flat. It turns out that free and projective are the same over a PID, and flat and torsion-free are the same over a PID (but free and projective is not the same as flat and torsion-free).

13.6. **Injective modules.** The following definition will probably only be used when R is a PID; otherwise I don't know if it is standard.

Definition 13.24. Let R be a commutative ring. An R-module M is said to be divisible if for all nonzerodivisors $a \in R$, the map $\times a : M \to M$ given by $m \mapsto am$ is surjective.

The strategy to show that R-Mod has enough injectives (R not necessarily commutative) will have the following two steps:

- When R is a (commutative) PID, an R-module is injective if and only if it is divisible. Moreover, in this case, any R-module can be embedded into an R-module that is divisible and hence injective.
- If every abelian group can be embedded in an injective abelian group (i.e., an injective $\mathbb{Z}$-module), then for more general (noncommutative, but associative and with 1) R, every R-module can be embedded in an injective R-module.

Today, we will carry out the second step; the first step will be discussed in Lecture 14.

Lemma 13.25. *Let $R \to S$ be a homomorphism of (not necessarily commutative) rings. Then $\mathrm{Hom}_R(S, -) : R\text{-}Mod \rightsquigarrow S\text{-}Mod$ (namely, the functor of "coextension of scalars from R to S" from Lecture 7) takes injective R-modules to injective S-modules (in other words, injectivity is preserved under coextension of scalars).*

Proof. In Lecture 7, we saw that coextension of scalars is right adjoint to restriction of scalars:

$$\mathrm{Hom}_S(-, \mathrm{Hom}_R(S, I)) \cong \mathrm{Hom}_R(\mathrm{Res}^S_R(-), I).$$

If I is an injective left R-module, then since $\mathrm{Res}^S_R(-)$ is exact and $\mathrm{Hom}_R(-, I)$ is exact, it follows that $\mathrm{Hom}_R(\mathrm{Res}^S_R(-), I)$ is exact, and hence so is $\mathrm{Hom}_S(-, \mathrm{Hom}_R(S, I))$. Thus, $\mathrm{Hom}_R(S, I)$ is an injective S-module. □

Remark 13.26. The above lemma might seem a bit "unintuitive", since coextension of scalars feels less intuitive than extension of scalars. It may help psychologically to notice that projectivity is preserved under extension of scalars.

The above lemma lets us finish step 2, since every ring R admits a ring homomorphism $\mathbb{Z} \to R$:

Lemma 13.27. *Let R be a (not necessarily commutative) ring, and M a left R-module. Suppose there exists an injective abelian group I containing the additive abelian group underlying M. Then there exists an injective R-module J containing the R-module M.*

Proof. We have injective homomorphisms of R-modules

$$M \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, M) \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, I),$$

where the latter morphism is composition with $M \hookrightarrow I$ and the former is given by $m \mapsto (r \mapsto rm)$. [42] We are done, since by Lemma 13.25, $J := \mathrm{Hom}_{\mathbb{Z}}(R, I)$ is an injective R-module. □

This completes step 2. We will start Lecture 14 with step 1.

[42] Check that this is indeed a homomorphism of R-modules, recalling the left R-module structure on $\mathrm{Hom}_{\mathbb{Z}}(R, M)$)

14. Lecture 14 — Derived functors

Today, unless otherwise stated (and we usually won't state otherwise), $\mathcal{A}$ and $\mathcal{B}$ will denote abelian categories, and any functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ will be assumed to be additive.

Blah blah. We will usually work with a left exact functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ and a right exact functor $G : \mathcal{A} \rightsquigarrow \mathcal{B}$. In Lecture 5, F was left adjoint and hence (under mild assumptions) right exact, and G was right adjoint and hence (under mild assumptions) left exact; so now the roles have switched. Further, we will switch another aspect of the notation: unlike the previous lecture, we will now mostly give proofs in the 'cochain complex/cohomology' situation and say that the 'chain complex/homology' situation is analogous. To avoid clumsiness of notation, rather than writing $A^\bullet_\circ \to A^\bullet \to A^\bullet_{\circ\circ}$ etc., we will start writing $A^\bullet_1 \to A^\bullet_2 \to A^\bullet_3$ etc.

14.1. **Continuation of the proof that R-Mod has enough injectives.** To prove that R-Mod has enough injectives, we reduced to the case where R was a PID. For this case, the key input is the following:

Proposition 14.1. *If R is a (commutative) PID, then an R-module M is injective if and only if it is divisible.*

Before proving Proposition 14.1, let us deduce from it that R-Mod has enough injectives.

Proposition 14.2. *For any (not necessarily commutative) ring R, R-Mod has enough injectives.*

Proof, assuming Proposition 14.1. In Lecture 13, we reduced this to the case where $R = \mathbb{Z}$ (see Lemma 13.27). Thus, we may now assume that R is a PID.

Suppose we can show the following: if M is an R-module and $m \in M$, then there exists a map $\varphi_m : M \to I_m$, with I a divisible (and hence, by Proposition 14.2, injective) module, and with $\varphi_m(m) \neq 0$. If this claim is granted, i.e., given such $\varphi_m : M \to I_m$ for all $m \in M$, $\prod_m \varphi_m : M \to \prod_m I_m$ is a monomorphism, and $\prod_m I_m$ is injective (since a product of injective modules is injective); therefore we will be done.

Thus, it remains to show the existence of such a $\varphi_m : M \to I_m$ for each $m \in M$. For this, it is enough to define an R-module injection $Rm \hookrightarrow I_m$ for some divisible R-module I_m: then, by the injectivity of I_m and the fact that $Rm \hookrightarrow M$ is a monomorphism, $Rm \hookrightarrow I_m$ extends to a map $\varphi_m : M \to I_m$, which does not vanish at m.

Note that $R/m \cong R/(d)$, for some $d \in R$. Let K be the quotient field of R: it is divisble as an R-module. If $d = 0$, we may take $\varphi_m : Rm \to I_m$ to be $Rm \xrightarrow{\cong} R \hookrightarrow K$, and we are done. If $d \neq 0$, we consider:

$$Rm \xrightarrow{\cong} R/(d) \xrightarrow{\cong} \frac{(1/d)R}{R} \hookrightarrow K/R,$$

and note that K/R too is divisible, and hence injective. □

Proof of Proposition 14.1. "⇒": If M is injective and $a \in R$ is a nonzerodivisor, apply $\mathrm{Hom}(-, M)$ to the monomorphism $\times a : R \to R$, to get that $\times a : M \to M$ is surjective.

"⇐": Suppose M is divisible. It is enough to show that any given injective homomorphism $M \hookrightarrow N$ of R-modules splits. Consider pairs (N', s'), where $N' \subset N$ is an R-module containing M, and $s' : N' \to M$ is a section to $M \hookrightarrow N'$. Given two such pairs (N', s') and (N'', s''), declare $(N', s') \leqslant (N'', s'')$ if $N \subset N''$, and $s''|_{N'} = s'$. The collection of such pairs is nonempty, since (M, id_M) belongs to this collection. Given a chain $\{(N_i, s_i)_i\}$ of such pairs, it is clear that $(N' = \bigcup_i N_i, s')$ is an upper bound for this collection, where $s' : N' \to M$ takes $n' \in N'$ to $s_i(n_i)$, for any i such that $n_i \in N_i$: this is independent of the choice of i, by the way we defined "$\leqslant$".

Thus, by Zorn's lemma, this collection has a maximal element, say (N', s'). Suppose $N' \neq N$, and let us get a contradiction. Choose $x \in N \backslash N'$, and set $N'' := N' + Rx \subset N$. We claim that we can extend the above section s' from N' to N''. Doing so is equivalent to extending $s'|_{N' \cap Rx} : N' \cap Rx \to M$ to an R-module homomorphism $Rx \to M$. Equivalently, consider the preimage I of $N' \cap Rx$ under $R \to Rx$ (given by $r \mapsto rx$). Then $I \subset R$ is an ideal, say $I = (a)$, and we have a composite homomorphism $\varphi : I \to N' \cap Rx \to M$. It is enough to extend $\varphi : I \to M$ to a homomorphism $\psi : R \to M$, for such a homomorphism will factor through $Rx \to M$ (since I contains $\mathrm{Ann}_R(x)$, as a consequence of the fact that $0 \in N' \cap Rx$), extending $N' \cap Rx \to M$.

This can be done since M is divisible, as follows: if $\varphi(a) = y \in M$, we have $\varphi(xa) = xy$ for all $x \in R$. Choose any $z \in M$ such that $y = az$, so $\varphi(xa) = xaz$ for all $x \in R$. Define $\psi(x) = xz$ for all $x \in R$. □

Exercise 14.3. (i) For the implication "⇐" in the above proof, only towards the end was it used that R is a PID. Use this to see that part of the above proof generalizes to the following important criterion:

Baer's criterion: Let R be a (not necessarily commutative) ring. Then a left R-module M is injective if and only if for all left ideals $I \subset R$, the restriction homomorphism $\mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(I, M)$ is surjective.

Note: Thus, instead of checking the defining criterion for the injectivity of M on all monomorphisms $N' \hookrightarrow N$, it is enough to consider the special cases consisting of injections $I \hookrightarrow R$, where $I \subset R$ is a left ideal.

(ii) Let R be a PID, $p \in R$ a prime element, and $n \in \mathbb{N}_{\geqslant 1}$. Show, using Baer's criterion or otherwise, that R/p^n is injective as an R/p^n-module.

(iii) Use the above exercise to prove the existence assertion in the structure theorem for finitely generated modules over a PID in the special case of torsion modules.

Note: Thus, to prove the existence assertion in the structure theorem, here is a sketch of a strategy, though we haven't given details for many of these steps:

- Show that finitely generated torsion-free modules over a PID R are free. Use this to reduce to the torsion case.
- Use the Chinese remainder theorem to show that $M = \bigoplus_p M[p^\infty]$, and thence reduce to the case where $\mathrm{Ann}_R(M) = (p^n)$ for some n.

- Use the above exercise to take care of this case (this will also involve 'reverse induction' on n).

14.2. **Injective and projective resolutions.** Recall that we wanted to define functors $\mathcal{A} \rightsquigarrow Kcoch(\mathcal{A})$, and $\mathcal{A} \rightsquigarrow Kch(\mathcal{A})$.

Definition 14.4. (i) A right resolution of $A \in \operatorname{Ob}\mathcal{A}$, or a resolution of A in $Coch(\mathcal{A})$, is an *exact* sequence:

$$0 \to A \to K^0 \to K^1 \to \dots$$

We will also abbreviate such a resolution to $A \to K^\bullet$. Sometimes, we will adopt a related but harmlessly conflicting terminology: we will say that $K^\bullet$ is a resolution of A, with the map $A \hookrightarrow K^0$ understood. Thus, a resolution of A can also be described as a sequence

$$K^\bullet : 0 \to K^0 \to K^1 \to \dots$$

that vanishes in degrees < 0 and is exact everywhere except at the 0-th place, but also given together with an identification $H^0(K^\bullet) \cong A$.

(ii) An injective resolution of A in $\mathcal{A}$ is a right resolution of A by injective objects.

(iii) A left resolution of $A \in \operatorname{Ob}\mathcal{A}$, or a resolution of A in $Ch(\mathcal{A})$, is an exact sequence

$$\dots \to L_1 \to L_0 \to A \to 0,$$

also written as $L_\bullet \to A$, or simply taken to be a sequence $L_\bullet \in Ch(\mathcal{A})$ vanishing in degrees < 0, exact everywhere except in degree 0, and considered along with an identification $H_0(L_\bullet) \cong A$.

(iv) A projective resolution of A is a left resolution of A by projective objects.

Lemma 14.5. *(i) If $\mathcal{A}$ has enough injectives, then every object of $\mathcal{A}$ has an injective resolution.*

(ii) If $\mathcal{A}$ has enough projectives, then every object of $\mathcal{A}$ has a projective resolution.

Proof. We will prove (i); the proof of (ii) is analogous. Given $A \in \operatorname{Ob}\mathcal{A}$, by the existence of enough injectives, we have a monomorphism $A \hookrightarrow I^0$ in $\mathcal{A}$, with I^0 injective. One similarly has a monomorphism $\operatorname{coker}(A \to I^0) \hookrightarrow I^1$, with I^1 injective. Note that

$$0 \to A \to I^0 \to I^1$$

is exact. Now repeat this with $\operatorname{coker}(I^0 \to I^1) \hookrightarrow I^2$, and continue. [43] □

Clearly, the injective or projective resolutions constructed in Lemma 14.5 are far from unique. As alluded to in Lecture 13, the best we can hope for is uniqueness up to homotopy equivalence, i.e., uniqueness up to a unique isomorphism in $Kcoch(\mathcal{A})$ or $Kch(\mathcal{A})$.

[43] If I understand Professor Nitin Nitsure right, induction doesn't give us infinite sequences, but only arbitrarily long sequences; one needs to combine this with a Zorn's lemma kind of argument to say that a sequence of infinite length exists.

Proposition 14.6. *(i) Suppose $A, B \in \operatorname{Ob}\mathcal{A}$, $A \to K^\bullet$ is a (not necessarily injective) resolution of A, and $B \to I^\bullet$ is an injective resolution of B. Then:*

(a) Any morphism $\alpha : A \to B$ lifts to a morphism $\alpha^\bullet : K^\bullet \to I^\bullet$: this means that $\alpha^\bullet$ is a morphism of cochain complexes, and $H^0(\alpha^\bullet) = \alpha : H^0(K^\bullet) = A \to B = H^0(I^\bullet)$.

(b) Any two lifts $\alpha^\bullet, \beta^\bullet$ of $\alpha : A \to B$ as above are homotopy equivalent, i.e., $\alpha^\bullet$ as in (a) is a unique morphism in $Kcoch(\mathcal{A})$.

(ii) Suppose $A, B \in \operatorname{Ob}\mathcal{A}$, $P_\bullet \to A$ is a projective resolution of A, and $L_\bullet \to B$ is a (not necessarily projective) resolution of B. Then:

(a) Any homomorphism $\alpha : A \to B$ lifts to a homomorphism $\alpha_\bullet : P_\bullet \to L_\bullet$: this means that $\alpha_\bullet$ is a morphism of chain complexes, and $H_0(\alpha_\bullet) = \alpha$.

(b) Any two lifts $\alpha_\bullet, \beta_\bullet$ of $\alpha : A \to B$ as above are homotopy equivalent, i.e., $\alpha_\bullet$ as in (a) is a unique morphism in $Kch(\mathcal{A})$.

Proof. We will prove (i); (ii) is analogous.

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & K^0 & \longrightarrow & K^1 & \longrightarrow & K^2 & \quad \dots . \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \alpha^0} & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_2} & \\ 0 & \longrightarrow & B & \longrightarrow & I^0 & \longrightarrow & I^1 & \longrightarrow & I^2 & \quad \dots \end{array}$$

By the injectivity of I^0, $A \to B \to I^0$ extends to some $\alpha^0 : K^0 \to I^0$ (this uses that $A \to K^0$ is a monomorphism). Note that $K^0 \to I^0 \to I^1$ vanishes on $\operatorname{im}(A \to K^0)$ (because $A \to K^0 \to I^0 \to I^1 = A \to B \to I^0 \to I^1 = 0$), and hence factors through a morphism $\operatorname{coker}(A \to K^0) \to I^1$. Since $\operatorname{coker}(A \to K^0) \hookrightarrow K^1$ is a monomorphism, the morphism $\operatorname{coker}(A \to K^0) \to I^1$ extends by the injectivity of I^1 to a morphism $\alpha^1 : K^1 \to I^1$. Now induct: in the next step one notes that $K^1 \to I^1 \to I^2$ factors through $\operatorname{coker}(K^0 \to K^1) \to I^2$, which extends to $\alpha^2 : K^2 \to I^2$, and so on. This proves (a).

Let us prove (b). Suppose we are given two liftings $\alpha^\bullet, \beta^\bullet$ of $\alpha : A \to B$. Then $\alpha^\bullet - \beta^\bullet$ lifts $0 : A \to B$. Thus, we may replace $\alpha^\bullet$ by $\alpha^\bullet - \beta^\bullet$, assume that $\alpha^\bullet$ lifts $0 : A \to B$, and show that $\alpha^\bullet$ is null homotopic. Consider

$$\begin{array}{ccccccc} \operatorname{coker}(A \to K^0) & \xhookrightarrow{d_K^0} & K^1 & \xrightarrow{d_K^1} & K^2 & . \\ \downarrow{\scriptstyle \alpha^0} & \swarrow{\scriptstyle s^1} & \downarrow{\scriptstyle \alpha^1} & \swarrow{\scriptstyle s^2} & \downarrow{\scriptstyle \alpha^2} & \\ I^0 & \xrightarrow[d_I^0]{} & I^1 & \xrightarrow[d_I^1]{} & I^2 & \end{array}$$

Since $\alpha = 0$, α^0 factors through $\operatorname{coker}(A \to K^0)$, explaining the left vertical arrow. Therefore, since $\operatorname{coker}(A \to K^0) \hookrightarrow K^1$ is a monomorphism and since I^0 is injective, we get $s^1 : K^1 \to I^0$ such that $\alpha^0 = s^1 \circ d_K^0$.

Now consider $\alpha^1 - d_I^0 \circ s^1$: since $(\alpha^1 - d_I^0 \circ s^1) \circ d_K^0 = d_I^0 \circ \alpha^0 - d_I^0 \circ s^1 \circ d_K^0 = 0$, so $\alpha^1 - d_I^0 \circ s^1$ factors thorough $K^1/\operatorname{im}(d_K^0)$. Since $K^1/\operatorname{im}(d_K^0) \hookrightarrow K^2$ is a monomorphism and since I^1 is

injective, this gives us $s^2 : K^2 \to I^1$ such that $s^2 \circ d_K^1 = \alpha^1 - d_I^0 \circ s^1$, i.e., $\alpha^1 = s^2 \circ d_K^1 + d_I^0 \circ s^1$. And so on. □

Corollary 14.7. *(i) Assume that A has enough injectives. Then there is a functor*

$$\mathrm{injres} : \mathcal{A} \rightsquigarrow Coch(\mathcal{A}),$$

sending each $A \in \mathrm{Ob}\,\mathcal{A}$ to some choice of injective resolution $I^\bullet$ of A, and any morphism $\alpha : A_1 \to A_2$ in $\mathcal{A}$ to the unique homotopy class of the liftings $\alpha^\bullet : I_1^\bullet \to I_2^\bullet$ lifting α (as given by Proposition 14.6(i)), where $I_1^\bullet$ and $I_2^\bullet$ are the 'chosen' injective resolutions of A_1 and A_2.

(ii) Assume that A has enough injectives. Then there is a functor

$$\mathrm{projres} : \mathcal{A} \rightsquigarrow Ch(\mathcal{A}),$$

sending each $A \in \mathrm{Ob}\,\mathcal{A}$ to some choice of projective resolution $P_\bullet$ of A, and any morphism $\alpha : A^1 \to A^2$ in $\mathcal{A}$ to the unique homotopy class of the liftings $\alpha_\bullet : P_\bullet^1 \to P_\bullet^2$ lifting α (as given by Proposition 14.6(ii)), where $P_\bullet^1$ and $P_\bullet^2$ are the 'chosen' projective resolutions of A^1 and A^2.

Proof. This is immediate from Proposition 14.6.

Here is a bit more of commentary. Given injective resolutions $I^\bullet$ and $J^\bullet$ for the same object $A \in \mathcal{A}$, (i) proposition gives us homotopy classes of morphisms $I^\bullet \to J^\bullet$ and $J^\bullet \to I^\bullet$. The compositions $I^\bullet \to J^\bullet$ and $J^\bullet \to I^\bullet$ are, by (i)(b), the homotopy classes of the identity morphisms $I^\bullet \to I^\bullet$ and $J^\bullet \to J^\bullet$. Similarly with projective resolutions. In other words, injective/projective resolutions, while not physically unique, are by Proposition 14.6 unique up to a unique isomorphism in $Kcoch(\mathcal{A})/Kch(\mathcal{A})$. One of course still needs the strong axiom of choice.

As explained in the statement of the corollary, the same proposition also allows us to define the functor at the level of morphisms. Compatibility with composition and identity morphisms is obvious. □

14.3. **The definition of derived functors.**

Notation 14.8. The following is probably non-standard notation, but I find it very notationally convenient. Please use this notation for understanding this lecture, but be circumspect in using it outside.

(i) Any (additive by the convention for this lecture) functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ determines an obvious functor $Coch(\mathcal{A}) \rightsquigarrow Coch(\mathcal{B})$, which will sometimes be abusively denoted by F itself. Thus, given

$$K^\bullet : \quad \ldots \overset{d^{i-1}}{\to} K^i \overset{d^i}{\to} K^{i+1} \overset{d^{i+1}}{\to} \ldots \quad \in \mathrm{Ob}\,Coch(\mathcal{A}),$$

we have

$$F(K^\bullet) : \quad \ldots \overset{F(d^{i-1})}{\to} F(K^i) \overset{F(d^i)}{\to} F(K^{i+1}) \overset{F(d^{i+1})}{\to} \ldots \quad \in \mathrm{Ob}\,Coch(\mathcal{B}).$$

Note that $F(K^\bullet)$ is indeed a complex, since $F(d^{i+1})\circ F(d^i) = F(d^{i+1}\circ d^i) = F(0) = 0$ for each i (F is additive by convention). It is immediate how this functor is defined at the level of morphisms.

(ii) If $\alpha^\bullet, \beta^\bullet : K_1^\bullet \to K_2^\bullet$ in $Coch(\mathcal{A})$ are homotopy equivalent, so that we have relations of the form $\alpha^i - \beta^i = s^{i+1}\circ d^i + d^{i-1}\circ s^i$, then the relations $F(\alpha^i) - F(\beta^i) = F(s^{i+1})\circ F(d^i) + F(d^{i-1})\circ F(s^i)$, a consequence of the additivity of F being assumed, imply that $F(\alpha^\bullet)$ and $F(\beta^\bullet)$ (as made sense of by (i) above) are homotopy equivalent. Thus, the functor $F : Coch(\mathcal{A}) \rightsquigarrow Coch(\mathcal{B})$ of (i) also desends to a functor $F : Kcoch(\mathcal{A}) \rightsquigarrow Kcoch(\mathcal{B})$ – note that we are denoting this also by F.

(iii) Similarly, any additive functor $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ determines functors $G : Ch(\mathcal{A}) \rightsquigarrow Ch(\mathcal{B})$ and $G : Kch(\mathcal{A}) \rightsquigarrow Kch(\mathcal{B})$.

Now we can define derived functors:

Definition 14.9. (i) Suppose $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is *left* exact, and assume that $\mathcal{A}$ has enough injectives. Then for all $i \in \mathbb{Z}$, the i-th *right* derived functor of F is the functor $R^iF : \mathcal{A} \rightsquigarrow \mathcal{B}$ defined as the composite

$$R^iF : \mathcal{A} \overset{\text{injres}}{\rightsquigarrow} Kcoch(\mathcal{A}) \overset{F}{\rightsquigarrow} Kcoch(\mathcal{B}) \overset{H^i}{\rightsquigarrow} \mathcal{B}$$

(Thus, for $i < 0$, $R^iF = 0$; note that we have used Notation 14.8).

(ii) Suppose $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is *right* exact, and assume that $\mathcal{A}$ has enough projectives. Then for all $i \in \mathbb{Z}$, the i-th *left* derived functor of G is the functor $L_iG : \mathcal{A} \rightsquigarrow \mathcal{B}$ defined as the composite

$$L_iG : \mathcal{A} \overset{\text{projres}}{\rightsquigarrow} Kch(\mathcal{A}) \overset{F}{\rightsquigarrow} Kch(\mathcal{B}) \overset{H_i}{\rightsquigarrow} \mathcal{B}$$

(Thus, for $i < 0$, $L_iF = 0$).

Remark 14.10. (i) Suppose $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact, and assume that $\mathcal{A}$ has enough injectives. Explicitly, the definition of $R^iF(A)$ above translates to the following: if $A \to (I^\bullet, d^\bullet)$ is an injective resolution of $A \in \operatorname{Ob}\mathcal{A}$, then

$$R^iF(A) \cong \frac{\ker(F(I^i) \overset{F(d^i)}{\to} F(I^{i+1}))}{\operatorname{im}(F(I^{i-1}) \overset{F(d^{i-1})}{\to} F(I^i))},$$

at the level of objects (its definition at the level of morphisms involves maps of complexes as justified using Proposition 14.6).

(ii) Further, we claim that $R^0F = F$. At the level of objects, we have

$$R^0F(A) = \frac{\ker(F(I^0) \to F(I^1))}{\operatorname{im}(F(I^{-1}) = 0 \overset{F(d^{-1})}{\to} F(I^i))} = \ker(F(I^0) \to F(I^1)) = F(A),$$

since $0 \to F(A) \to F(I^0) \to F(I^1)$ is exact by the left exactness of F. It is an easy exercise to see this holds at the level of morphisms as well (because of the 'lifting' of morphisms being in Proposition 14.6).

Note that this agrees with our plans: we wanted to extend F to a cohomological δ-functor, which will take the form $(\{R^iF\}_i, \{\delta^i\})$.

(iii) Similarly, suppose $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is right exact, and that $\mathcal{A}$ has enough projectives. If $(P_\bullet, \partial_\bullet) \to A$ is a projective resolution of $A \in \operatorname{Ob} \mathcal{A}$, then

$$L_i G(A) \cong \frac{\ker(G(P_i) \to G(P_{i-1}))}{\operatorname{im}(G(P_{i+1}) \to G(P_i))},$$

at the level of objects (its definition at the level of morphisms involves maps of complexes as justified using Proposition 14.6).

(iv) As with the identity $R^0 F = F$, it is easy to see that $L_0 G = G$.

14.4. **The horseshoe lemma.** We need one more proposition to extend $\{R^i F\}_i$ to a cohomological δ-functor $(\{R^i F\}_i, \{\delta^i\}_i)$, and $\{L_i G\}_i$ to a homological δ-functor $(\{L_i G\}_i, \{\delta_i\}_i)$: the following, which is sometimes called the horseshoe lemma (due to the shape of the diagram in equation (64) below, with I_2^0 removed: or rather, one can view the lemma as choosing injective resolutions of A_1 and A_3, and filling in a compatible injective resolution of A_2).

Proposition 14.11. *(i) Suppose $\mathcal{A}$ has enough injectives. If $0 \to A_1 \to A_2 \to A_3 \to 0$ is a short exact sequence in $\mathcal{A}$, then there exist injective resolutions $I_1^\bullet, I_2^\bullet$ and $I_3^\bullet$ of A_1, A_2 and A_3, respectively, and a short exact sequence*

$$0 \to I_1^\bullet \to I_2^\bullet \to I_3^\bullet \to 0$$

in $Coch(\mathcal{A})$ lifting $0 \to A_1 \to A_2 \to A_3 \to 0$. [44]

(ii) Suppose $\mathcal{A}$ has enough projectives. If $0 \to A^1 \to A^2 \to A^3 \to 0$ is a short exact sequence in $\mathcal{A}$, then there exist projective resolutions $P_\bullet^1, P_\bullet^2$ and $P_\bullet^3$ of A^1, A^2 and A^3, respectively, and a short exact sequence

$$0 \to P_\bullet^1 \to P_\bullet^2 \to P_\bullet^3 \to 0$$

in $Ch(\mathcal{A})$ lifting $0 \to A^1 \to A^2 \to A^3 \to 0$.

Proof. We will prove (i); (ii) is analogous. Though the proof is simple, or rather because the proof is simple, we will write it out in detail.

First we claim that it is enough to obtain a commutative diagram of the form:

(64)
$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \longrightarrow & A_2 & \longrightarrow & A_3 & \longrightarrow & 0\,, \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & I_1^0 & \longrightarrow & I_2^0 & \longrightarrow & I_3^0 & \longrightarrow & 0 \end{array}$$

where the top row is the given exact sequence, the bottom row is a short exact sequence with I_1^0, I_2^0 and I_3^0 injective, and where the vertical arrows are all monomorphisms.

[44] i.e., $I_1^\bullet \to I_2^\bullet$ lifts $A_1 \to A_2$ in the sense mentioned in Proposition 14.6(i), and similarly $I_2^\bullet \to I_3^\bullet$ lifts $A_2 \to A_3$.

Indeed, if we construct such a diagram, then by the snake lemma and the fact that the vertical arrow $A_3 \hookrightarrow I_3^0$ is a monomorphism, we would get an exact sequence at the level of cokernels:

$$0 \to \operatorname{coker}(A_1 \hookrightarrow I_1^0) \to \operatorname{coker}(A_2 \hookrightarrow I_2^0) \to \operatorname{coker}(A_3 \hookrightarrow I_3^0) \to 0.$$

Recalling that injective resolutions of each A_i are constructed by the inductive procedure described in Lemma 14.5(i), it is easy to iterate this procedure and get $0 \to I_1^\bullet \to I_2^\bullet \to I_3^\bullet \to 0$ as desired.

Thus, let us construct (64). Choose monomorphisms $u : A_1 \hookrightarrow I_1^0$ and $w : A_3 \hookrightarrow I_3^0$, and set $I_2^0 := I_1^0 \oplus I_3^0$, giving us all of the following diagram except its middle vertical arrow:

(65)
$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \longrightarrow & A_2 & \longrightarrow & A_3 & \longrightarrow & 0\ . \\ & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle v:=(u',w')} & & \downarrow{\scriptstyle w} & & \\ 0 & \longrightarrow & I_1^0 & \longrightarrow & I_2^0 = I_1^0 \oplus I_3^0 & \longrightarrow & I_3^0 & \longrightarrow & 0 \end{array}$$

Note that I_2^0 is injective, and the bottom row (whose maps are the obvious ones) is exact. We wish to define a monomorphism $v : A_2 \to I_2^0$ such that the entire diagram commutes.

Defining $v : A_2 \to I_2^0 = I_1^0 \oplus I_3^0$ is equivalent to defining $u' : A_2 \to I_1^0$ and $w' : A_2 \to I_3^0$: [45]

- Define $u' : A_2 \to I_1^0$ to be an extension of $A_1 \to I_1^0$ to a map $A_2 \to I_1^0$: such an extension exists since I_1^0 is injective, and since $A_1 \hookrightarrow A_2$ is a monomorphism.
- Define w' to be $A_2 \to A_3 \to I_3^0$.

Thus, we get $v := (v', w') : A_2 \to I_2^0$, as shown in (65). We now handle the remaining assertions to complete the proof of the proposition:

- $A_2 \to I_2^0$ is a monomorphism: indeed, $\ker v = \ker(u', w') \hookrightarrow A_2$ factors through $\ker(w')$, is just $A_1 \hookrightarrow A_2$. Thus, $\ker v \hookrightarrow A_2$ identifies with the kernel of $u' \circ (A_1 \hookrightarrow A_2) = u$, which is 0.
- The left square commutes: this is because $w' \circ (A_1 \hookrightarrow A_2)$ is 0.
- The right square commutes: this is because $A_2 \hookrightarrow I_2^0 = I_1^0 \oplus I_3^0 \to I_3^0$ is w', which by definition equals $A_2 \to A_3 \to I_3^0$.

□

Let us also record a comment that Rajesh made during the lecture:

Lemma 14.12. *Let $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ be left exact and $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ right exact.*

(i) If $\mathcal{A}$ has enough injectives, then for any injective object $I \in \operatorname{Ob} \mathcal{A}$,

$$R^i F(I) = \begin{cases} F(I), & \text{if } i = 0, \text{ and} \\ 0, & \text{if } i > 0. \end{cases}$$

[45] We are using notation and language used for *R-Mod*, but note that this makes sense for an abelian category: I_2^0 is a product of I_1^0 and I_3^0, etc.

(ii) If $\mathcal{A}$ has enough projectives, then for any projective object $P \in \operatorname{Ob}\mathcal{A}$,

$$L_iG(P) = \begin{cases} G(P), & \text{if } i = 0, \text{ and} \\ 0, & \text{if } i > 0. \end{cases}$$

Proof. This is because an injective resolution of I can be taken to be $0 \to I \to 0 \to 0 \to \dots$, and a projective resolution of P can be taken to be $\cdots \to 0 \to \cdots \to 0 \to P \to 0$. □

14.5. **Derived functors are delta functors.** Now we can prove that left/right derived functors indeed give homological/cohomological δ-functors; but for that, one of the smaller steps is a general enough fact to be separately given as an exercise:

Exercise 14.13. (i) Given an exact sequence $0 \to A' \to A \to A'' \to 0$, show that the following are equivalent:
- (a) The sequence is split exact, i.e., we can identify A with $A' \oplus A''$ in such a way that $A' \to A = A' \oplus A''$ and $A' \oplus A'' = A \to A''$ become the obvious maps.
- (b) The map $A' \to A$ is a split monomorphism, i.e., there exists $s : A \to A'$ such that $A' \to A \xrightarrow{s} A' = \mathrm{id}_{A'}$.
- (c) The map $A \to A''$ is a split epimorphism, i.e., there exists $s' : A'' \to A$ such that $A'' \xrightarrow{s'} A \to A'' = \mathrm{id}_{A''}$.

(ii) Show that any additive functor $\mathcal{A} \rightsquigarrow \mathcal{B}$ takes a split exact sequence to a split exact sequence.

Theorem 14.14. *(i) Assume that $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact, and that $\mathcal{A}$ has enough injectives. Then the $R^iF, i \geqslant 0$, are part of a cohomological δ-functor, $(\{R^iF\}_i, \{\delta^i\}_i)$ (which automatically extends F in the obvious sense, i.e., $R^0F = F$).*

(ii) Assume that $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is right exact, and that $\mathcal{A}$ has enough projectives. Then the $L_iF, i \geqslant 0$, are part of a homological δ-functor $(\{L_iF\}_i, \{\delta_i\}_i)$ (which automatically extends G in the obvious sense, i.e., $L_0G = G$).

Proof. We will prove (i); (ii) is analogous.

Given a short exact sequence $0 \to A_1 \to A_2 \to A_3 \to 0$, we choose a short exact sequence $0 \to I_1^\bullet \xrightarrow{f^\bullet} I_2^\bullet \xrightarrow{g^\bullet} I_3^\bullet \to 0$ of injective resolutions as in Proposition 14.11.

We claim that

$$0 \to F(I_1^\bullet) \xrightarrow{F(f^\bullet)} F(I_2^\bullet) \xrightarrow{F(g^\bullet)} F(I_3^\bullet) \to 0$$

is exact, where the notation $F(I_j^\bullet)$ is as in Notation 14.8. For this, it is enough to show that each $0 \to F(I_1^i) \to F(I_2^i) \to F(I_3^i) \to 0$ is exact.

To see this, note that $0 \to I_1^i \to I_2^i \to I_3^i \to 0$ is split exact, as a consequence of I_1^i being injective and $0 \to I_1^i \to I_2^i$ being a monomorphism (giving an $s : I_2^i \to I_1^i$ as in Exercise 14.13(i) above). Thus, by Exercise 14.13(ii), $0 \to F(I_1^i) \to F(I_2^i) \to F(I_3^i) \to 0$ is split exact, and hence exact. This proves the claim.

By the first proposition of Lecture 13, namely Proposition 13.3, we get a long exact sequence:

$$0 \to H^0(F(I_1^\bullet)) \overset{H^0(F(f^\bullet))}{\to} H^0(F(I_2^\bullet)) \overset{H^0(F(g^\bullet))}{\to} H^0(F(I_3^\bullet)) \overset{\delta^0}{\to} H^1(F(I_1^\bullet)) \overset{H^1(F(f^\bullet))}{\to} \dots$$
$$\dots \overset{\delta^{i-1}}{\to} H^i(F(I_1^\bullet)) \overset{H^i(F(f^\bullet))}{\to} H^i(F(I_2^\bullet)) \overset{H^i(F(g^\bullet))}{\to} H^i(F(I_3^\bullet)) \overset{\delta^i}{\to} H^{i+1}(F(I_1^\bullet)) \overset{H^{i+1}(F(f^\bullet))}{\to} \dots$$

This defines the δ^i (one for each short exact sequence $0 \to A_1 \to A_2 \to A_3 \to 0$) that will be part of the cohomological δ-functor $(\{R^iF\}_i, \{\delta^i\}_i)$.

By the definition of derived functors (see Definition 14.9), this is the same as:

$$0 \to R^0F(A_1) \overset{R^0F(f)}{\to} R^0F(A_2) \overset{R^0F(g)}{\to} R^0F(A_3) \overset{\delta^0}{\to} R^1F(A_1) \to \dots$$
$$\dots \overset{\delta^{i-1}}{\to} R^iF(A_1) \overset{R^iF(f)}{\to} R^iF(A_2) \overset{R^iF(g)}{\to} R^iF(A_3) \overset{\delta^i}{\to} R^{i+1}F(A_1) \overset{R^{i+1}F(f)}{\to} \dots .$$

(the δ^i's are just borrowed from the previous equation).

We actually haven't yet proved that the δ^i's are functorial, which is needed to complete the proof that $(\{R^iF\}_i, \{\delta^i\}_i)$ is a cohomological δ-functor. I haven't carefully checked the following, so be especially careful with what follows.

By the functoriality assertion for the δ^n of the first proposition of Lecture 13, namely Proposition 13.3, this follows if we prove that given a morphism between short exact sequences $0 \to A_1 \to A_2 \to A_3 \to 0$ and $0 \to B_1 \to B_2 \to B_3 \to 0$, we have a morphism of short exact sequences of complexes

$$\begin{array}{ccccccccc} 0 & \longrightarrow & I_1^\bullet & \longrightarrow & I_2^\bullet & \longrightarrow & I_3^\bullet & \longrightarrow & 0\ , \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & J_1^\bullet & \longrightarrow & J_2^\bullet & \longrightarrow & J_3^\bullet & \longrightarrow & 0 \end{array}$$

where the top row is associated to $0 \to A_1 \to A_2 \to A_3 \to 0$ as in Proposition 14.11, the bottom row is associated similarly to $0 \to B_1 \to B_2 \to B_3 \to 0$, and each $I_i^\bullet \to J_i^\bullet$ lifts $A_i \to B_i$ in the sense mentioned in Proposition 14.6.

We try to ensure this by modifying the construction of Proposition 14.11, associated to the two given short exact sequences. As in that proposition, it is enough to do the first step. We first construct $I_1^0, I_3^0, J_1^0, J_3^0$, but we additionally impose the following, as we may: $A_1 \hookrightarrow I_1^0$ is actually a composite $A_1 \hookrightarrow A_2 \hookrightarrow I_1^0$ of monomorphisms.

We can clearly ensure that $I_1^0 \to J_1^0$ and $I_3^0 \to J_3^0$ lift $A_1 \to B_1$ and $A_2 \to B_2$ (see the proof of Proposition 14.6(i)(a)). Examining the construction of $A_2 \to I_2^0 = I_1^0 \oplus I_3^0$ and $B_2 \to J_2^0 = J_1^0 \oplus J_3^0$ found in Proposition 14.6, the only thing needed to ensure that we can choose a lifting $I_2^0 \to J_2^0$ of $A_2 \to B_2$ that both restricts to $I_1^0 \to J_1^0$ and induces $I_3^0 \to J_3^0$, is to ensure that the maps $A_2 \to I_1^0$ and $B_2 \to J_1^0$ as in that construction can be chosen to

fit into a commutative diagram:

$$\begin{array}{ccc} A_2 & \longrightarrow & B_2 \\ \downarrow & & \downarrow \\ I_1^0 & \longrightarrow & J_1^0 \end{array} .$$

By the injectivity of J_1^0, this can be ensured if we can ensure that $A_2 \to I_1^0$ is a monomorphism: but this is okay since we have chosen $A_1 \hookrightarrow I_1^0$ to be a composite $A_1 \hookrightarrow A_2 \hookrightarrow I_1^0$. □

Remark 14.15. The R^iF and the L^iG are not merely cohomological and homological δ-functors, but also 'universal' cohomological and homological δ-functors, in a sense that I hope to discuss in Lecture 15.

14.6. **The** Ext **and the** Tor **functors.**

Definition 14.16. Let R be a (not necessarily commutative) ring.

(i) Let $A \in \mathrm{Ob}\, \mathit{Mod}\text{-}R$ be a right R-module, and let $B \in \mathrm{Ob}\, R\text{-}\mathit{Mod}$ be a left R-module. We define, for $i \geqslant 0$:
  (a) $\mathrm{Tor}_i^R(A, -) := L_i(A \otimes_R -) : R\text{-}\mathit{Mod} \rightsquigarrow \mathit{AbGrp}$.
  (b) $\mathrm{tor}_i^R(-, B) := L_i(- \otimes_R B) : \mathit{Mod}\text{-}R \rightsquigarrow \mathit{AbGrp}$.
  These definitions make sense since we have shown that $R\text{-}\mathit{Mod}$ and (similarly) $\mathit{Mod}\text{-}R$ have enough projectives.

  When R is commutative, these functors may and shall be viewed as valued in $R\text{-}\mathit{Mod}$ instead of in AbGrp.

(ii) Let $A, B \in \mathrm{Ob}\, R\text{-}\mathit{Mod}$ be left R-modules. We define, for $i \geqslant 0$:
  (a) $\mathrm{Ext}^i_R(A, -) := R^i(\mathrm{Hom}_R(A, -)) : R\text{-}\mathit{Mod} \rightsquigarrow \mathit{AbGrp}$.
  (b) $\mathrm{ext}^i_R(-, B) := R^i(\mathrm{Hom}_R(-, B)) : R\text{-}\mathit{Mod}^{op} \rightsquigarrow \mathit{AbGrp}$.
  These definitions make sense since we have shown that $R\text{-}\mathit{Mod}$ has enough injectives, and so does $R\text{-}\mathit{Mod}^{op}$, because injective objects of $R\text{-}\mathit{Mod}^{op}$ are simply the projective objects of $R\text{-}\mathit{Mod}$.

  Again, when R is commutative, these functors may and shall be viewed as valued in $R\text{-}\mathit{Mod}$ instead of in AbGrp.

Remark 14.17. (i) The above is nonstandard notation: one knows that tor = Tor and ext = Ext in a suitable sense, so we will soon be writing only Tor and Ext.

(ii) Note that, in keeping with our conventions, the 'i' of tor_i^R and Tor_i^R are subscripts, while the 'i' of Ext^i_R and ext^i_R are superscripts. When i is subscripted R is superscripted, and the other way round.

(iii) Thus, if $P_\bullet$ is a projective resolution of A in $\mathit{Mod}\text{-}R$ and $Q_\bullet$ one of B in $R\text{-}\mathit{Mod}$, then we have

$$\mathrm{Tor}_i^R(A, B) = \frac{\ker(A \otimes_R Q_i \to A \otimes_R Q_{i-1})}{\mathrm{im}(A \otimes_R Q_{i+1} \to A \otimes_R Q_i)}, \qquad \mathrm{tor}_i^R(A, B) = \frac{\ker(P_i \otimes_R B \to P_{i-1} \otimes_R B)}{\mathrm{im}(P_{i+1} \otimes_R B \to P_i \otimes_R B)}.$$

(iv) On the other hand, if $P_\bullet$ is a projective resolution of A in R-Mod, or in other words an injective resolution of A in R-Mod^{op}, and if $I^\bullet$ is an injective resolution of B in R-Mod, we have:

$$\mathrm{Ext}^i_R(A,B) = \frac{\ker(\mathrm{Hom}(A,I^i) \to \mathrm{Hom}(A,I^{i+1}))}{\mathrm{im}(\mathrm{Hom}(A,I^{i-1}) \to \mathrm{Hom}(A,I^i))}, \quad \mathrm{ext}^i_R(A,B) = \frac{\ker(\mathrm{Hom}(P_i,B) \to \mathrm{Hom}(P_{i+1},B))}{\mathrm{im}(\mathrm{Hom}(P_{i-1},B) \to \mathrm{Hom}(P_i,B))}.$$

These formulas can sometimes help calculate these objects. We will see some examples below.

(v) If $0 \to N' \to N \to N'' \to 0$ is an exact sequence of left R-modules, then by Theorem 14.14 we get, for any right R-module M, an exact sequence of abelian groups:

$$\cdots \to \mathrm{Tor}^R_i(M,N') \to \mathrm{Tor}^R_i(M,N) \to \mathrm{Tor}^R_i(M,N'') \xrightarrow{\delta_i} \mathrm{Tor}^R_{i-1}(M,N') \to \ldots$$
$$\cdots \to \mathrm{Tor}^R_1(M,N'') \xrightarrow{\delta_1} M \otimes_R N' \to M \otimes_R N \to M \otimes_R N'' \to 0.$$

A similar comment applies with tor, with M', M, M'' of an exact sequence occuring in the first argument (and hence, by Theorem 14.18 quoted below, with Tor too). Often, this sequence too can be used to compute values of Tor.

(vi) If $0 \to N' \to N \to N'' \to 0$ is an exact sequence of left R-modules, then by Theorem 14.14 we get, for any left R-module M, an exact sequence of abelian groups:

$$0 \to \mathrm{Hom}_R(M,N') \to \mathrm{Hom}_R(M,N) \to \mathrm{Hom}_R(M,N'') \to \mathrm{Ext}^1_R(M,N') \to \ldots$$
$$\cdots \to \mathrm{Ext}^i_R(M,N'') \xrightarrow{\delta^i} \mathrm{Ext}^{i+1}_R(M,N') \to \mathrm{Ext}^{i+1}_R(M,N) \to \mathrm{Ext}^{i+1}_R(M,N'') \to \ldots.$$

A similar comment applies with ext, and with M', M, M'' occuring in the second argument (and hence, by Theorem 14.18 quoted below, with Ext too). Often, this sequence too can be used to compute values of Ext.

In the next lecture, I hope to sketch a proof of the following:

Theorem 14.18. *Let R be a (not necessarily commutative) ring.*

(i) For all $A \in \mathrm{Ob}\, Mod\text{-}R$ and $B \in \mathrm{Ob}\, R\text{-}Mod$, and for all $i \geqslant 0$, we have an isomorphism $\mathrm{Tor}^R_i(A,B) \cong \mathrm{tor}^R_i(A,B)$, functorial in A and B, and specializing when $i = 0$ to the identity map $A \otimes_R B \to A \otimes_R B$.

(ii) For all $A, B \in \mathrm{Ob}\, R\text{-}Mod$, and for all $i \geqslant 0$, we have an isomorphism $\mathrm{Ext}^i_R(A,B) \cong \mathrm{ext}^i_R(A,B)$, functorial in A and B, and specializing when $i = 0$ to the identity map $\mathrm{Hom}_R(A,B) \to \mathrm{Hom}_R(A,B)$.

When R is commutative, these are identifications of R-modules.

Some of the examples below will use the following exercise:

Exercise 14.19. Work out or look up a proof of the following: If R is a (commutative) PID, show that any submodule of a free R-module is free.

In the following examples, we are distinguishing between Tor and tor, and between Ext and ext; note however that they are eventually the same by Theorem 14.18 above, so these examples give more information than what is literally written below.

Example 14.20. (i) Let R be a commutative ring, and let $a \in R$ be a nonzero divisor. We claim that:

$$\operatorname{Tor}_i^R(M, R/(a)) \cong \begin{cases} M/aM, & \text{if } i = 0, \\ M[a] := \{m \in M \mid am = 0\}, & \text{if } i = 1, \text{ and} \\ 0, & \text{otherwise.} \end{cases} \tag{66}$$

Indeed, this is because a projective resolution of $R/(a)$ can be taken to be $0 \to R \xrightarrow{\times a} R \to 0$, which on tensoring with M becomes

$$0 \to M \xrightarrow{\times a} M \to 0,$$

whose homology is as claimed. Now applying the long exact sequence for Tor (see Remark 14.17(v)), we get an exact sequence:

$$0 \to M'[a] \to M[a] \to M''[a] \to M'/aM' \to M/aM \to M''/aM'' \to 0,$$

which was also the topic of Exercise 12.30 from Lecture 12.

$\operatorname{Tor}_1(M, R/a)$ being the a-torsion of M gives some idea of why the notation 'Tor' is used.

(ii) A similar computation shows that if R is a commutative ring and $a \in R$ is a nonzerodivisor, then:

$$\operatorname{Ext}^i_R(R/a, M) = \begin{cases} M[a], & \text{if } i = 0, \\ M/aM, & \text{if } i = 1, \text{ and} \\ 0, & \text{if } i > 1. \end{cases}$$

(iii) If R is a PID, then for all $i \geqslant 2$, we have $\operatorname{Tor}_i^R(M, N) = \operatorname{tor}_i^R(M, N) = 0$. This follows from the fact that, by Exercise 14.19 above, any module over a PID has a free resolution $P_\bullet$, with $P_i = 0$ for $i > 1$.

(iv) A similar reasoning gives that if R is a PID, then for all $i \geqslant 2$, $\operatorname{ext}^i_R(M, N) = 0$. Without appealing to Theorem 14.18 quoted above, we can also see directly that for all $i \geqslant 2$, $\operatorname{Ext}^i_R(M, N) = 0$ (with R a PID): this is because if $N \hookrightarrow I^0$ with I^0 injective and hence divisible, then $I^1 := I^0/N$ is automatically divisible and hence injective; therefore $0 \to I^0 \to I^1 \to 0 \to 0 \to \dots$ is an injective resolution of N.

(v) Let $R = k[x]/(x^n)$. Think of k as an R-module, where x acts as 0. We claim that $\operatorname{ext}^i_R(k, k) = k$ for all $i \geqslant 0$. For this, note that we have the following projective resolution for the R-module k:

$$\dots \xrightarrow{\cdot x^{n-1}} R \xrightarrow{\cdot x} R \xrightarrow{\cdot x^{n-1}} R \xrightarrow{\cdot x} \dots \xrightarrow{\cdot x^{n-1}} R \xrightarrow{\cdot x} R \to k \to 0.$$

Now apply $\operatorname{Hom}(-k)$, to get:

$$0 \to k \xrightarrow{0} k \xrightarrow{0} k \xrightarrow{0} \dots \xrightarrow{0} k \xrightarrow{0} \dots$$

Exercise 14.21. Let R be a commutative ring, and $I, J \subset R$ ideals. Show that:

$$\mathrm{Tor}_1^R(R/I, R/J) \cong \frac{I \cap J}{IJ}.$$

Thus, to belabour a painfully obvious point, $\mathrm{Tor}_1^R(R/I, R/J)$ measures how far $I \cap J$ fails to be IJ.

15. Lecture 15 — Acyclic objects, universal δ-functors, more on Ext and Tor

Unless otherwise stated, $\mathcal{A}$ and $\mathcal{B}$ will be abelian categories. Any functors $F, G : \mathcal{A} \rightsquigarrow \mathcal{B}$ will automatically be assumed to be additive; F will often be left exact, while G will often be right exact.

15.1. **A long exact sequence in terms of short exact sequences.** Before proceeding, let us recall/make an observation frequently seen:

Remark 15.1. (i) Any long exact sequence can be broken up into a chain of short exact sequences: if $0 \to K^1 \to K^2 \to \dots$ is a long exact sequence, then setting $C^i = \operatorname{coker}(K^{i-1} \to K^i)$ for each i, we get short exact sequences

$$0 \to C^0 := K^0 \to K^1 \to C^1 \to 0,\ 0 \to C^1 \to K^2 \to C^2 \to 0, \dots .$$

(ii) Conversely, if we are given short exact sequences $0 \to C^0 := K^0 \to K^1 \to C^1 \to 0$, $0 \to C^1 \to K^2 \to C^2 \to 0, \dots$, then we get by splicing together a long exact sequence

$$0 \to K^0 \to K^1 \to \dots,$$

where each $K^i \to K^{i+1}$ is defined to be $K^i \to C^i \to K^{i+1}$.

There are 'chain analogues' of the above, of exact sequences $\dots \to K^1 \to K^0 \to 0$.

15.2. **Dimension shifting.** The following definition will be ad hoc now; it is partially motivated by Proposition 15.6 below, and further motivated later into this lecture.

Definition 15.2. (i) An additive functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ between abelian categories is said to be effaceable or erasable, if for all $A \in \operatorname{Ob}\mathcal{A}$ there exists a monomorphism $A \hookrightarrow K$, for some $K \in \operatorname{Ob}\mathcal{A}$ such that $F(K) = 0$. The condition $F(K) = 0$ may be expressed by saying that K erases F. We may also say that $A \hookrightarrow K$ effaces or erases A.

(ii) Let $T = (\{T^n\}_n, \{\delta^n\}_n)$ be a cohomological δ-functor between $\mathcal{A}$ and $\mathcal{B}$. We say that T is effaceable, if T^i is effaceable for each $i \geqslant 1$ (but not necessarily for $i = 0$).

(iii) An additive functor $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ between abelian categories is said to be coeffaceable or coerasable, if for all $A \in \operatorname{Ob}\mathcal{A}$ there exists an epimorphism $L \twoheadrightarrow A$, for some $L \in \operatorname{Ob}\mathcal{A}$ such that $G(L) = 0$. The condition $G(L) = 0$ may be expressed by saying that L erases G. We may also say that $L \twoheadrightarrow A$ effaces or erases A.

(iv) A homological δ-functor $(\{T_n\}_n, \{\delta_n\}_n)$ is said to be coeffaceable or coerasable if T_n is coeffaceable for each $i \geqslant 1$ (but not necessarily for $i = 0$).

Notation 15.3. In Lecture 14, associated to a left exact functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ such that all the right derived functors R^nF exist (as per our definition from Lecture 14, that is – thus, this existence is equivalent to $\mathcal{A}$ having enough injectives), we saw a construction of a cohomological δ-functor $(\{R^nF\}_n, \{\delta^n\}_n)$. We will refer to this δ-functor $(\{R^nF\}_n, \{\delta^n\}_n)$ as the cohomological δ-functor obtained by right deriving F (warning: this terminology

may be nonstandard). Similarly, we will talk of homological δ-functors $(\{L_nG\}_n, \{\delta_n\}_n)$ obtained by left deriving G.

Example 15.4. (i) Any δ-functor $(\{R^nF\}_n, \{\delta^n\}_n)$ obtained by right deriving $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is effaceable: [46] this is because given each $A \in \mathrm{Ob}\,\mathcal{A}$ we have a monomorphism $A \hookrightarrow I$ with I injective, and we noticed in Lecture 14 that R^nF vanishes on injective objects for each $n \geqslant 1$.

(ii) Similarly, δ-functors $(\{L_nG\}_n, \{\delta_n\}_n)$ obtained by left deriving right exact functors $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ are coeffaceable.

Remark 15.5. To illustrate dimension shifting, consider a cohomological δ-functor $(\{T^n\}_n, \{\delta^n\}_n)$, together with an exact sequence:

$$0 \to A \to K \to C \to 0,$$

If K effaces F^n for all $n \geqslant 1$, then we have for all $n \geqslant 1$ (though not necessarily for $n = 0$), an exact sequence:

$$0 = T^n(K) \to T^n(C) \to T^{n+1}(A) \to T^{n+1}(K) = 0,$$

yielding, for each $n \geqslant 1$:

$$T^n(C) \cong T^{n+1}(A). \tag{67}$$

Thus, we will often be able to reduce the proofs of results about $T^{n+1}(A)$ to those about $T^n(C)$, provided $n \geqslant 1$.

Thus, one might expect to similarly reduce results about $T^{n+r}(A)$ to those about $T^n(C)$ for a different C, and that is what the following proposition says:

Proposition 15.6. *(i) Let $(\{T^n\}_n, \{\delta^n\}_n)$ be a cohomological δ-functor from $\mathcal{A}$ to $\mathcal{B}$. Let*

$$0 \to A \to K^0 \to K^1 \to K^2 \to \cdots \to K^{r-1} \to C \to 0 \tag{68}$$

be an exact sequence in $\mathcal{A}$, where $T^n(K^i) = 0$ for each i and each $n \geqslant 1$. Then for all $n \geqslant 1$ we have

$$T^n(C) \cong T^{n+r}(A).$$

(ii) Let $(\{T_n\}_n, \{\delta_n\}_n)$ be a homological δ-functor from $\mathcal{A}$ to $\mathcal{B}$. Let

$$0 \to C \to L_{r-1} \to \cdots \to L_1 \to L_0 \to A \to 0 \tag{69}$$

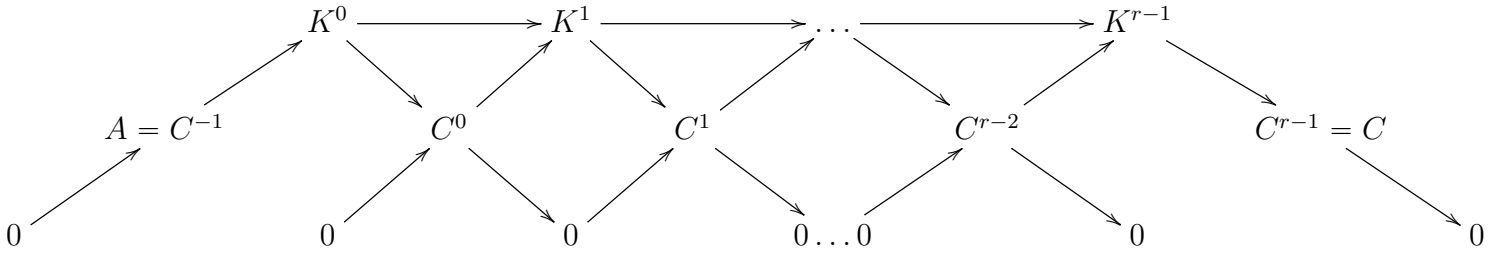
be an exact sequence in $\mathcal{A}$, where $T_n(L_i) = 0$ for each i, and for each $n \geqslant 1$. Then for all $n \geqslant 1$ we have

$$T_n(C) \cong T_{n+r}(A).$$

[46] Thus, we are implicitly assuming here that F is left exact and that $\mathcal{A}$ has enough injectives.

Proof. We will prove (i); (ii) is analogous.

The proof is just an iterated application of the above discussion (see around (67)) together with breaking up the given exact sequence into short exact sequences as in Remark 15.1(i), and as represented by the following diagram:
(70)



Applying the discussion of Remark 15.5 to the exact sequences $0 \to C^{i-1} \to K^i \to C^i$, as i ranges between 0 and r, we get for each $n \geqslant 1$:

$$T^n(C) = T^n(C^{r-1}) = T^{n+1}(C^{r-2}) = T^{n+2}(C^{r-3}) = \cdots = T^{n+r}(C^{-1}) = T^{n+r}(A),$$

as desired. □

We close this subsection with some definitions that the above considerations lead to; we may not use them much in this course but they seem useful even to just understand this kind of mathematics:

Definition 15.7. (i) Given an exact sequence

$$0 \to A \to I^1 \to I^2 \to \cdots \to I^{r-1} \to C \to 0,$$

with each I^i injective, we call C an r-th cosyzygy of A. Thus, by Proposition 15.6(i), whenever $\mathcal{A}$ has enough injectives and $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact, then for any r-th cosyzygy C of A we have $R^{n+r}F(A) = R^nF(C)$ for all $n \geqslant 1$ (each I^i being injective automatically satisfies $R^nF(I) = 0$ for all $n \geqslant 1$).

(ii) Given an exact sequence

$$0 \to C \to P_{r-1} \to \cdots \to P_1 \to P_0 \to A \to 0,$$

with each P_i projective, we call C an r-th syzygy of A. Thus, by Proposition 15.6(ii), whenever $\mathcal{A}$ has enough projectives and $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is right exact, then for any r-th syzygy C of A we have $L_{n+r}G(A) = L_nG(C)$ for all $n \geqslant 1$ (each P_i being projective is automatically satisfies $L_nG(P_i) = 0$ for all $n \geqslant 1$).

(iii) Let $T = (\{T^n\}_n, \{\delta^n\}_n)$ be a cohomological δ-functor from $\mathcal{A}$ to $\mathcal{B}$. We say that $A \in \operatorname{Ob}\mathcal{A}$ has T-dimension $\leqslant d$, if $T^n(A) = 0$ for all $n \geqslant d+1$. If the right derived functors of a left exact functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ exist, by F-dimension we mean $(\{R^nF\}_n, \{\delta^n\}_n)$-dimension.

(iv) Let $T = (\{T_n\}_n, \{\delta_n\}_n)$ be a homological δ-functor from $\mathcal{A}$ to $\mathcal{B}$. We say that $A \in \operatorname{Ob}\mathcal{A}$ has T-dimension $\leqslant d$, if $T_n(A) = 0$ for all $n \geqslant d+1$. If the left derived

functors of a right exact functor $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ exist, by G-dimension we mean $(\{L_nG\}_n, \{\delta_n\}_n)$-dimension.

15.3. **Acyclic objects.** The following definitions are closely related to those of the previous subsection:

Definition 15.8. (i) Let $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ be a left exact functor between abelian categories, and assume that $\mathcal{A}$ has enough injectives. An object $A \in \operatorname{Ob}\mathcal{A}$ is said to be F-acyclic, if $R^iF(A) = 0$ for all $i \geqslant 1$ (i.e., if A has F-dimension $\leqslant 0$).

(ii) Relatedly, if $T = (\{T^n\}_n, \{\delta^n\}_n)$ is a cohomological δ-functor from an abelian category $\mathcal{A}$ to an abelian category $\mathcal{B}$, an object $A \in \operatorname{Ob}\mathcal{A}$ is said to be T-acyclic, if $T^i(A) = 0$ for all $i \geqslant 1$ (i.e., if A has T-dimension $\leqslant 0$).

Thus, in the situation of (i), A is F-acyclic if and only if it is $(\{R^nF\}_n, \{\delta^n\}_n)$-acyclic, where $(\{R^nF\}_n, \{\delta^n\}_n)$ is the cohomological δ-functor obtained by right deriving F.

(iii) Similarly, let $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ be a right exact functor between abelian categories, and assume that $\mathcal{A}$ has enough projectives. An object $A \in \operatorname{Ob}\mathcal{A}$ is said to be G-acyclic, if $L_iG(A) = 0$ for all $i \geqslant 1$ (i.e., if A has G-dimension $\leqslant 0$).

(iv) Relatedly, if $T = (\{T_n\}_n, \{\delta_n\}_n)$ is a homological δ-functor from an abelian category $\mathcal{A}$ to an abelian category $\mathcal{B}$, an object $A \in \operatorname{Ob}\mathcal{A}$ is said to be T-acyclic, if $T_i(A) = 0$ for all $i \geqslant 1$ (i.e., if A has T-dimension $\leqslant 0$).

Thus, in the situation of (iii), A is G-acyclic if and only if it is $(\{L_nG\}_n, \{\delta_n\}_n)$-acyclic, where $(\{L_nG\}_n, \{\delta_n\}_n)$ is the homological δ-functor obtained by left deriving G.

(v) If $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact and $\mathcal{A}$ has enough injectives, a right resolution $A \to K^\bullet$ of $A \in \operatorname{Ob}\mathcal{A}$ is said to be F-acyclic, if each K^i is F-acyclic.

(vi) Similarly, if $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is right exact and A has enough projectives, a left resolution $L_\bullet \to A$ is said to be G-acyclic, if each L_i is G-acyclic.

(vii) There is a related notion of 'acyclic', without any 'F-acyclic', which seems to be a distinct notion. Namely:

We will often also refer to a cochain complex $K^\bullet$, where $K^i = 0$ for all $i < 0$, as acyclic if $H^i(K^\bullet) = 0$ for all $i > 0$. Thus, it is exact everywhere except at 0, but it may not be exact, because $H^0(K^\bullet)$ is allowed to be nonzero. Thus, if $A \to K^\bullet$ is a right resolution of A, then $K^\bullet$ is not exact (unless $A = 0$), but it is acyclic.

(viii) Similarly we define an acyclic chain complex. If $L_\bullet \to A$ is a left resolution of A, then $L_\bullet$ is not exact (unless $A = 0$), but it is acyclic.

Remark 15.9. Thus, F-dimension measures how far T is from being F-acyclic, etc.

For various definitions we will encounter along the above lines, there will be three special cases involving functors that we have studied so far: an 'injective version', a 'projective version' and a 'flat version': the first two refer to $\operatorname{Hom}(A, -)$ and $\operatorname{Hom}(-, B)$, while the third will be restricted to module categories, and will involve tensor products.

Example 15.10. Let $\mathcal{A}$ be an abelian category.

(i) Assume that $\mathcal{A}$ has enough injectives. If $I \in \operatorname{Ob}\mathcal{A}$ is injective, then we have seen in Lecture 14 that for any left-exact functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$, we have $R^iF(I) = 0$ for all $i \geqslant 1$. i.e., I is F-acyclic for any left exact functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$. Conversely, we will see in Exercise 15.25 below that any $I \in \operatorname{Ob} A$ that is acyclic for each $\operatorname{Hom}(A, -)$, as A varies over $\operatorname{Ob}\mathcal{A}$, is injective.

(ii) Assume that $\mathcal{A}$ has enough projectives. Projective objects of $\mathcal{A}$ are the same as injective objects of $\mathcal{A}^{op}$. Thus, (i) translates to the following: any projective object $P \in \operatorname{Ob}\mathcal{A}$ is F-acyclic for any left exact functor $F : \mathcal{A}^{op} \to \mathcal{B}$, and conversely, (we will see in Exercise 15.25 below that) any $P \in \operatorname{Ob} A$ that is acyclic for each $\operatorname{Hom}_{\mathcal{A}^{op}}(A, -) = \operatorname{Hom}_{\mathcal{A}}(-, A)$, as A varies over $\operatorname{Ob}\mathcal{A}$, is projective.

(iii) If $N \in \operatorname{Ob} R\text{-}Mod$ is flat, then we will (essentially) see that N is $(M \otimes_R -)$-acyclic for any $M \in \operatorname{Ob} Mod\text{-}R$ (similarly with $N \in \operatorname{Ob} Mod\text{-}R$ and the $- \otimes_R M$ with $M \in \operatorname{Ob} R\text{-}Mod$). Rather, the details in the commutative case are left to Exercise 15.25 below, and the general case is similar.

15.4. **Derived functors can be computed using acyclic resolutions.** Now let us see that derived functors of F can be computed using F-acyclic resolutions, which can be more convenient in some situations than injective resolutions:

Theorem 15.11. *(i) Suppose $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact, and that $\mathcal{A}$ has enough injectives. If $A \to K^\bullet$ is an F-acyclic resolution of A, then $R^iF(A)$ can be computed as $H^i(F(K^\bullet))$ for each i. More precisely, if $K^\bullet \to I^\bullet$ is any lift of $\mathrm{id}_A : A \to A$ (see Proposition 14.6 from Lecture 14), then the induced map*

$$H^i(F(K^\bullet)) \to H^i(F(I^\bullet)) \overset{\text{recall}}{=} R^iF(A)$$

is an isomorphism.

(ii) Suppose $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is right exact, and that $\mathcal{A}$ has enough projectives . If $L_\bullet \to A$ is a G-acyclic resolution of A, then $L_iG(A)$ can be computed as $H_i(G(L_\bullet))$ for each i. More precisely, if $P^\bullet \to L^\bullet$ is any lift of $\mathrm{id}_A : A \to A$ (see Proposition 14.6 from Lecture 14), then the induced map

$$L_iG(A) \overset{\text{recall}}{=} H_i(G(P_\bullet)) \to H_i(G(L_\bullet))$$

is an isomorphism.

The proof will use:

Lemma 15.12. *If $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact and $Y^\bullet : 0 \to Y^0 \to Y^1 \to \dots$ is an exact cochain complex with each Y^i F-acyclic, then $0 \to F(Y^0) \to F(Y^1) \to \cdots$ is exact. A similar assertion applies to right exact $G : \mathcal{A} \rightsquigarrow \mathcal{B}$, and chain complexes $\cdots Y_1 \to Y_0 \to 0$.*

This lemma will in turn be a simple consequence of two simple observations, the first of which is breaking up a long exact sequence into short exact sequences (Remark 15.1 above). The second simple observation is the following lemma.

Lemma 15.13. *Let $0 \to A \to B \to C \to 0$ be an exact sequence in $\mathcal{A}$.*

(i) If $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is left exact, $\mathcal{A}$ has enough injectives, and A and B are F-cyclic, then C is F-acyclic as well.
(ii) If $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is right exact, $\mathcal{A}$ has enough projectives, and B and C are G-acyclic, then A is G-acyclic as well.

Proof. We will prove the first assertion; the second assertion is analogous. For each $i \geqslant 1$, the long exact sequence obtained by applying F to $0 \to A \to B \to C \to 0$ gives that for all $i \geqslant 1$:

$$0 = R^iF(B) \to R^iF(C) \xrightarrow{\delta} R^{i+1}F(A) = 0$$

is exact, so $R^iF(C) = 0$. □

Proof of Lemma 15.12. We will prove the assertion concerning F; the assertion concerning G is analogous. By Remark 15.1(i), we get short exact sequences $0 \to C^i \to Y^{i+1} \to C^{i+1} \to 0$ for each $i \geqslant 0$, where $C^i = \operatorname{coker}(Y^{i-1} \to Y^i)$ (thus, $C^0 = Y^0$). By part (ii) of the same remark, it is enough to show that each $0 \to F(C^i) \to F(Y^{i+1}) \to F(C^{i+1}) \to 0$ is exact.

By the long exact sequence obtained by applying F to $0 \to C^i \to Y^{i+1} \to C^{i+1} \to 0$, this follows if we show that $R^1F(C^{i+1}) = 0$ for each $i \geqslant 0$. This is because, by Lemma 15.13, $C^{i+1} = \operatorname{coker}(Y^i \to Y^{i+1})$ is F-acyclic for each $i \geqslant 0$. □

Proof of Theorem 15.11. We will prove (i); (ii) is analogous.

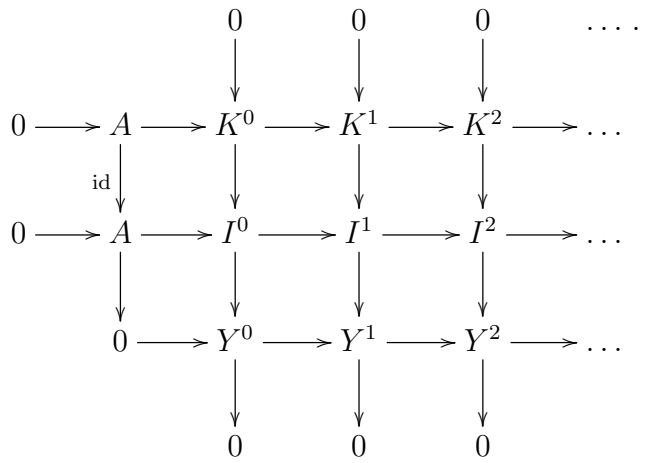
Step 1. Choosing $I^\bullet$. We have the freedom to choose $I^\bullet$. We claim that we can ensure that $K^i \to I^i$ is a monomorphism for each $i \geqslant 0$. These is a standard 'pushout + enough injectives' argument, which we now recall. For this, first choose the monomorphism $A \hookrightarrow I^0$ to be of the form $A \to K^0 \hookrightarrow I^0$. Letting $C^0 = \operatorname{coker}(A \to K^0) = \operatorname{im}(K^0 \to K^1) \hookrightarrow K^1$ and $D^0 = \operatorname{coker}(A \to I^0)$, form the pushout, the rightmost square in the following diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & K^0 & \longrightarrow & C^0 & \longrightarrow & K^1 \\ & & \| & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & A & \longrightarrow & I^0 & \longrightarrow & D^0 & \longrightarrow & E^1 \hookrightarrow I^1 \end{array} .$$

We choose $I^0 \to I^1$ to be $I^0 \to D^0 \to E^1 \hookrightarrow I^1$, where $E^1 \hookrightarrow I^1$ is a monomorphism of E^1 into an injective object I^1 of $\mathcal{A}$. Now since pushouts preserve monomorphisms (see Corollary 12.23(ii) from Lecture 12), $D^0 \to E^1$ is a monomorphism, which implies that $A \to I^0 \to I^1$ is exact. Now induct: in the next step, one considers $\operatorname{coker}(K^0 \to K^1)$ and $\operatorname{coker}(I^0 \to I^1)$ in place of C^0 and D^0. Clearly, $K^\bullet \to I^\bullet$ then lifts $\mathrm{id}_A : A \to A$, and consists of monomorphisms $K^i \to I^i$.

Step 2. The complex $Y^\bullet$ of acyclics. Let $Y^i = \operatorname{coker}(K^i \to I^i)$, for each $i \geqslant 0$. By Lemma 15.13, we get that each Y^i is F-acyclic (use that each K^i is F-acyclic, and so is each I^i; each injective object has been observed to be acyclic for each left exact functor). Moreover, each $I^i \to I^{i+1}$ induces $Y^i \to Y^{i+1}$.

Now we have a commutative diagram with exact rows and columns:



Applying F, we get a sequence of chain complexes

$$0 \to F(K^\bullet) \to F(I^\bullet) \to F(Y^\bullet) \to 0$$

(here $K^\bullet$ is the complex $0 \to K^0 \to K^1 \to \dots$, and similarly with $I^\bullet$ and $Y^\bullet$). This complex is exact by Lemma 15.12, since each Y^i is F-acyclic.

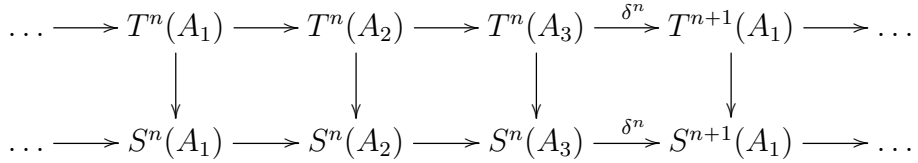
The long exact sequence for cohomology associated to this sequence contains, for $i \geqslant 1$, an exact subsequence

$$H^{i-1}(F(Y^\bullet)) \to H^i(F(K^\bullet)) \to H^i(F(I^\bullet)) \to H^i(F(Y^\bullet)),$$

where the map $H^i(F(K^\bullet)) \to H^i(F(I^\bullet))$ is indeed as in the statement of the theorem. Since $H^{i-1}(F(Y^\bullet)) = H^i(F(Y^\bullet)) = 0$ by Lemma 15.12, this forces the map $H^i(F(K^\bullet)) \to H^i(F(I^\bullet)) = R^i(F^\bullet)$ to be an isomorphism, as desired. □

15.5. **Universal δ-functors.** We now study the notion of universality for homological and cohomological δ-functors, which (as we will see below) are very useful in relating various derived functors to each other.

Definition 15.14. (i) A morphism between cohomological δ-functors $(\{T^n\}_n, \{\delta^n\}_n)$ and $(\{S^n\}_n, \{\delta^n\}_n)$ is a unique sequence of natural transformations $\{T^n \to S^n\}_{n \geqslant 1}$ such that for each exact sequence $0 \to A_1 \to A_2 \to A_3 \to 0$ in $\mathcal{A}$, and each $n \in \mathbb{N}$, the square involving the δ^n's in the following diagram commutes:



Note that this condition in fact gives the commutativity of the the entire diagram above, because the $T^n \to S^n$ are natural transformations.

(ii) Similarly, we define a morphism of homological δ-functors.

(iii) A cohomological δ-functor $(\{T^n\}_n, \{\delta^n\}_n)$ is said to be universal if, given any cohomological δ-functor $(\{S^n\}_n, \{\delta^n\}_n)$ and a natural transformation $T^0 \to S^0$, there exists a unique morphism of δ-functors $(\{T^n\}_n, \{\delta^n\}_n) \to (\{S^n\}_n, \{\delta^n\}_n)$ extending, in an obvious sense, the given natural transformation $T^0 \to S^0$.

(iv) A homological δ-functor $(\{T_n\}_n, \{\delta_n\}_n)$ is said to be universal if, given any homological δ-functor $(\{S_n\}_n, \{\delta_n\}_n)$ and a natural transformation $S_0 \to T_0$, there exists a unique morphism of δ-functors $(\{S_n\}_n, \{\delta_n\}_n) \to (\{T_n\}_n, \{\delta_n\}_n)$ extending, in an obvious sense, the given natural transformation $S_0 \to T_0$.

Theorem 15.15. *(i) Any cohomological δ-functor $(\{R^nF\}_n, \{\delta^n\}_n)$ obtained by right deriving a left exact functor $F : \mathcal{A} \rightsquigarrow \mathcal{B}$ is universal.*

(ii) Any homological δ-functor $(\{L_nG\}_n, \{\delta_n\}_n)$ obtained by left deriving a right exact functor $G : \mathcal{A} \rightsquigarrow \mathcal{B}$ is universal.

Corollary 15.16. *(i) If $\alpha : F_1 \to F_2$ is a natural transformation between left exact functors $\mathcal{A} \rightsquigarrow \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are abelian categories and $\mathcal{A}$ has enough injectives, then it extends to a unique morphism of cohomological δ-functors $f^\alpha : (\{R^nF_1\}_n, \{\delta^n\}_n) \to (\{R^nF_2\}_n, \{\delta^n\}_n)$. Moreover, this respects composition: $f_{\alpha\circ\beta} = f_\alpha \circ f_\beta$ (when applicable).*

(ii) Similarly, if $\alpha : G_1 \to G_2$ is a natural transformation between right exact functors $\mathcal{A} \rightsquigarrow \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are abelian categories and $\mathcal{A}$ has enough projectives, then it extends uniquely to a morphism of homological δ-functors $f_\alpha : (\{L_nG_1\}_n, \{\delta_n\}_n) \to (\{L_nG_2\}_n, \{\delta_n\}_n)$. Moreover, this respects composition: $f_{\alpha\circ\beta} = f_\alpha \circ f_\beta$.

Proof. The existence and the uniqueness of the f_α are immediate from Theorem 15.15. The uniqueness gives that the f_α respect composition. □

Remark 15.17. It is easy to see to describe the morphisms $R^nF_1 \to R^nF_2$ given by the above corollary more concretely: for any $A \in \operatorname{Ob}\mathcal{A}$, say with an injective resolution $I^\bullet$, the natural transformation $F_1 \to F_2$ gives rise to a morphism of complexes $\eta =$ "$\alpha(I^\bullet)$" $: F_1(I^\bullet) \to F_2(I^\bullet)$, and $R^nF_1(A) \to R^nF_2(A)$ is given by:

$$H^n(\eta) : R^nF_1(A) = H^n(F_1(I^\bullet)) \to H^n(F_2(I^\bullet)) = R^nF_2(A).$$

Indeed, these morphisms commute with the "δ" (use the functoriality of the long exact sequence associated to a short exact sequence of complexes), and hence give morphisms of δ-functors, which by the above corollary is unique.

Similarly with the morphisms $L_nG_1 \to L_nG_2$.

Since the $(\{R^nF\}_n, \{\delta^n\}_n)$ are effaceable and the $(\{L_nG\}_n, \{\delta_n\}_n)$ are coeffaceable (see Example 15.4), Theorem 15.15 follows from the following theorem:

Theorem 15.18 (Grothendieck, published in Tohoku)**.** *(i) Any effaceable cohomological δ-functor $(\{T^n\}_n, \{\delta^n\}_n)$ is universal.*

(ii) Any coeffaceable homological δ-functor $(\{T_n\}_n, \{\delta_n\}_n)$ is universal.

Proof. The following is much more detailed than what was discussed in the lecture. We will prove (i); the proof of(ii) is analogous. Assume that these functors are from an abelian category $\mathcal{A}$ to an abelian category $\mathcal{B}$.

We are given a natural transformation $f^0 : T^0 \to S^0$; we need to define natural transformations $f^n : T^n \to S^n$ for each $n \geqslant 1$, such that $(f^n)_{n\geqslant 0}$ is a morphism of δ-functors $(\{T^n\}_n, \{\delta^n\}_n) \to (\{S^n\}_n, \{\delta^n\}_n)$.

Step 1: A construction of $T^1(A) \to S^1(A)$. The natural transformation $f^1 : T^1 \to S^1$ is a collection of maps $f^1 : T^1(A) \to S^1(A)$, as A varies over $\operatorname{Ob}\mathcal{A}$. The first step is to give a construction for $f^1 : T^1(A) \to S^1(A)$, for a fixed $A \in \operatorname{Ob}\mathcal{A}$.

We are assuming the effaceability of T^n for each $n \geqslant 1$. Use the effaceability of T^1 to erase A with a monomorphism $u : A \hookrightarrow K$; thus, $T^1(K) = 0$. Let $C = \operatorname{coker}(A \to K)$, so we have an exact sequence $0 \to A \xrightarrow{u} K \to C \to 0$.

Thus, we get a commutative diagram with exact rows:

$$\begin{array}{ccccccc} T^0(K) & \longrightarrow & T^0(C) & \xrightarrow{\delta_T^1} & T^1(A) & \longrightarrow & T^1(K) = 0\ . \\ \downarrow{\scriptstyle f^0} & & \downarrow{\scriptstyle f^0} & & \vdots{\scriptstyle ?} & & \\ S^0(K) & \longrightarrow & S^0(C) & \xrightarrow{\delta_S^1} & S^1(A) & & \end{array}$$

To prove the existence of the dotted arrow, which will be our $f^1 : T^1(A) \to S^1(A)$, thanks to $T^0(C) \to T^1(A)$ being an epimorphism, it is enough to show that $\ker(\delta_T^1) \subset \ker(\delta_S^1 \circ f^0)$. [47] By the exactness of the top row, this follows if we show that the composite $T^0(K) \to T^0(C) \to S^0(C) \to S^1(A)$ is zero, which is the case since it is also the composite $T^0(K) \to S^0(K) \to S^0(C) \to S^1(A)$. This defines, depending on $u : A \hookrightarrow K$, $f^1 : T^1(A) \to S^1(A)$ uniquely (uniquely because $T^0(C) \to T^1(A)$ is an epimorphism).

Step 2: Well-definedness of $f^1 : T^1(A) \to S^1(A)$. Now we show that $f^1 : T^1(A) \to S^1(A)$ is independent of the choice of $u : A \hookrightarrow K$. Suppose $u : A \hookrightarrow K$ and $u' : A \hookrightarrow K'$ are monomorphisms, with $T^1(K) = T^1(K') = 0$.

First assume that u' factors as $A \xrightarrow{u} K \hookrightarrow K'$: though this is a special case, but we will reduce the general case to this case. $K \to K'$ induces a map $w : C \to C'$, where $C = \operatorname{coker} u$ and $C' = \operatorname{coker} u'$. We then get a diagram

$$\begin{array}{ccccccc} T^0(C) & \xrightarrow{T^0(w)} & T^0(C') & \xrightarrow{\delta_T} & T^1(A) & \longrightarrow & 0\ , \\ \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle f_1} & & \\ S^0(C) & \xrightarrow{S^0(w)} & S^0(C') & \xrightarrow{\delta_S} & S^1(A) & & \end{array}$$

[47] more explanation: in our abelian category context, this should of course be interpreted as saying that $\ker \delta_T^1 \hookrightarrow T^0(C)$ factors through $\ker(\delta_S^1 \circ f^0) \hookrightarrow T^0(C)$, i.e., $\ker \delta_T^1 \hookrightarrow T^0(C)$ has trivial composite with $\delta_S^1 \circ f^0$; this will give that $\delta_S^1 \circ f^0$ factors through the cokernel $T^0(C) \to T^1(A)$ of $\ker \delta_T^1 \hookrightarrow T^0(C)$.

where the right vertical arrow is defined using u': this gives the commutativity of the right square, while the commutativity of the left square follows from $T^0 \to S^0$ being a natural transformation. Therefore, the outer square commutes. The horizontal arrows of the outer square are $\delta_T : T^0(C) \to T^1(A)$ and $\delta_S : S^0(C) \to S^1(A)$, simply because $(\{T^n\}_n, \{\delta^n\}_n)$ and $(\{S^n\}_n, \{\delta^n\}_n)$ are δ-functors. It follows that the right vertical arrow is also the map $T^1(A) \to S^1(A)$ defined by u. Thus, u and u' define the same map $T^1(A) \to S^1(A)$, as needed.

Now consider the general case, where u' does not factor as $A \xrightarrow{u} K \hookrightarrow K'$. To reduce this case to the case treated earlier, it is enough to show that there exists a monomorphism $u'' : A \hookrightarrow K''$, through which both u and u' factor, and where $T^1(K'') = 0$: by the already treated special case, this will give that the maps $T^1(A) \to S^1(A)$ defined using u and u' each coincide with the map $T^1(A) \to S^1(A)$ defined using u'', and hence agree with each other (again, this is quite a standard mode of reduction in mathematics: to show that two choices give the same result, one reduces to the case where there is an inclusion or some such relation between the choices). To do this, we will use the following standard push-out argument (this is similar to an argument we saw in the proof of Lemma 15.12; we present it separately to invoke it for other parts of this proof later).

A push-out argument: Consider

$$\begin{array}{ccc} A & \xrightarrow{u} & K \\ {\scriptstyle u'}\downarrow & & \downarrow \\ K' & \longrightarrow & K \coprod_A K' \end{array} \hookrightarrow K'' \quad ,$$

where we have, using the effaceability of T^1, chosen an object $K'' \in \operatorname{Ob} \mathcal{A}$ and a monomorphism $K \coprod_A K' \hookrightarrow K''$ such that $T^1(K'') = 0$. To finish the proof that u and u' both factor through a common monomorphism $A \hookrightarrow K''$ and hence yield the same morphism $T^1(A) \to S^1(A)$, it is enough to prove that $K \to K \coprod_A K'$ and $K' \to K \coprod_A K'$ are monomorphisms.

But since the top horizontal arrow and the left vertical arrow are monomorphisms, this follows from (applying once to each of these arrows) the fact that pushouts preserve monomorphisms (see Corollary 12.23(ii) from Lecture 12). This completes the proof that $f^1 : T^1(A) \to S^1(A)$ is well-defined.

Step 3. The functoriality of $T^1(A) \to S^1(A)$ in A. So far, we have only defined $f^1 : T^1(A) \to S^1(A)$ for each $A \in \operatorname{Ob} \mathcal{A}$. We need to show that f^1 is a natural transformation, i.e., that for each $w : A \to B$ in $\mathcal{A}$, the following diagram commutes:

$$\begin{array}{ccc} T^1(A) & \xrightarrow{T^1(w)} & T^1(B) \\ {\scriptstyle f^1}\downarrow & & \downarrow{\scriptstyle f^1} \\ S^1(A) & \xrightarrow{S^1(w)} & S^1(B) \end{array} .$$

We will choose compatible erasing monomorphisms $u : A \to K$ and $v : B \to L$, for which we form a pushout diagram, the first diagram below:

$$\begin{array}{ccc} A & \xhookrightarrow{u} & K \\ {\scriptstyle w}\downarrow & & \downarrow \\ B & \longrightarrow & B \coprod_A K \end{array} \hookrightarrow L \qquad \Rightarrow \qquad \begin{array}{ccc} A & \xhookrightarrow{u} & K \\ {\scriptstyle w}\downarrow & & \downarrow{\scriptstyle t} \\ B & \xhookrightarrow{v} & L \end{array},$$

where we have, using the effaceability of T^1, chosen an object $L \in \mathrm{Ob}\,\mathcal{A}$ and a monomorphism $B \coprod_A K \hookrightarrow L$ such that $T^1(L) = 0$. Then $v : B \to L$ is a monomorphism, as shown in the second square above, since $u : A \to K$ is a monomorphism, and because pushouts preserve monomorphisms as quoted earlier. Let $C = \mathrm{coker}(u : A \to K)$, and $D = \mathrm{coker}(v : B \to L)$, so $K \to L$ induces a map $s : C \to D$.

Now consider the diagram:

$$\begin{array}{ccccc} T^0(C) & & \xrightarrow{T^0(s)} & & T^0(D) \\ & {\scriptstyle \delta_T}\searrow & & \swarrow{\scriptstyle \delta_T} & \\ & T^1(A) & \xrightarrow{T^1(w)} & T^1(B) & \\ {\scriptstyle f^0}\downarrow & {\scriptstyle f^1}\downarrow & & \downarrow{\scriptstyle f^1} & \downarrow{\scriptstyle f^0} \\ & S^1(A) & \xrightarrow{S^1(w)} & S^1(B) & \\ & {\scriptstyle \delta_S}\nearrow & & \nwarrow{\scriptstyle \delta_S} & \\ S^0(C) & & \xrightarrow[S^0(s)]{} & & S^0(D) \end{array}.$$

To finish the proof of the functoriality of $f^1 : T^1 \to S^1$, in the above diagram, it is enough to prove the commutativity of the inner square. The outer square commutes, since T^0 is already given as a natural transformation. The top and the bottom trapezia commute since $(\{T^n\}_n, \{\delta^n\}_n)$ and $(\{S^n\}_n, \{\delta^n\}_n)$ are δ-functors. f^1 was defined so as to make the side trapezia commute. In other words, everything other than the inner square is known to commute.

This much is not enough to force that the inner square commutes, but (as we will see below) it turns out to be forced by the additional constraint that $\delta_T : T^0(C) \to T^1(A)$ in the diagram is epimorphism (recall that this epimorphism-ness was seen while constructing $f^1 : T^1(A) \to S^1(A)$, and followed from $T^1(K) = T^1(L) = 0$).

Since $\delta_T : T^0(C) \to T^1(A)$ is an epimorphism, it is enough to show that $T^0(C) \to T^1(A) \to T^1(B) \to S^1(B)$ equals $T^0(C) \to T^1(A) \to S^1(A) \to S^1(B)$. We proceed using the known

commutativity of the quadrilaterals other than the inner square, as follows:

$$\begin{aligned}(T^0(C) \to T^1(A) \to S^1(A) \to S^1(B)) &= (T^0(C) \to S^0(C) \to S^1(A) \to S^1(B)) \\ = (T^0(C) \to S^0(C) \to S^0(D) \to S^1(B)) &= (T^0(C) \to T^0(D) \to S^0(D) \to S^1(B)). \\ = (T^0(C) \to T^0(D) \to T^1(B) \to S^1(B)) &= (T^0(C) \to T^1(A) \to T^1(B) \to S^1(B))\end{aligned}$$

This finishes the proof that f^1 is indeed a natural transformation $T^1 \to S^1$.

Step 4. The commutativity with δ. To get a morphism of δ-functors, for each exact sequence

$$0 \to A_1 \to A_2 \to A_3 \to 0$$

in $\mathcal{A}$, we need to establish the commutativity of

$$\begin{array}{ccc} T^0(A_3) & \xrightarrow{\delta_T} & T^1(A_1) \\ {\scriptstyle f^0}\downarrow & & \downarrow{\scriptstyle f^1} \\ S^0(A_3) & \xrightarrow{\delta_S} & S^1(A_1) \end{array} . \tag{71}$$

The proof of this will be similar to part of the proof that $f^1 : T^1(A) \to S^1(A)$ is well-defined.

We choose a monomorphism $v : A_2 \to K$ such that $F(K) = 0$. Then $A_1 \to A_2 \xrightarrow{v} K$ is a monomorphism, and can be used to compute $T^1(A_1) \to S^1(A_1)$. We have a map $w : A_3 = \operatorname{coker}(A_1 \to A_2) \to \operatorname{coker}(A_1 \to K) =: C$, so we have a commutative diagram with exact rows,

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \longrightarrow & A_2 & \longrightarrow & A_3 & \longrightarrow & 0 \\ & & \| & & \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle w} & & \\ 0 & \longrightarrow & A_1 & \longrightarrow & K & \longrightarrow & C & \longrightarrow & 0 \end{array} .$$

We now consider the diagram

$$\begin{array}{ccccc} T^0(A_3) & \xrightarrow{T^0(w)} & T^0(C) & \xrightarrow{\delta_T} & T^1(A_1) \\ {\scriptstyle f^0}\downarrow & & \downarrow{\scriptstyle f^0} & & \downarrow{\scriptstyle f^1} \\ S^0(A_3) & \xrightarrow{S^0(w)} & S^0(C) & \xrightarrow{\delta_S} & S^1(A_1) \end{array} .$$

The right square is commutative because $f^1 : T^1(A^1) \to S^1(A^1)$ was defined so as to ensure it. The left square is commutative because $f^0 : T^0 \to S^0$ is a natural transformation. Therefore, the outer rectangle is commutative, and it suffices to see that the outer rectangle is the same as the square (71).

The vertical arrows of both the outer rectangle and the square are already the same. That the horizontal arrows match follows from the fact that $(\{T^n\}, \{\delta^n\}_n)$ and $(\{S^n\}, \{\delta^n\}_n)$ are δ-functors.

Thus, we have defined $f^1 : T^1 \to S^1$ using $f^0 : T^0 \to S^0$, and shown that it has all the required properties. Once we have defined $f^n : T^n \to S^n$ as a natural transformation and

verified its commutativity with δ, a similar argument lets us define $f^{n+1} : T^{n+1} \to S^{n+1}$ and verify its commutativity with δ (we did not use the left exactness of T^0 or S^0 anywhere in the proof). This concludes the proof. □

Exercise 15.19. Formulate and prove an assertion to the effect that the morphism $\{f^n\}_n : (\{T^n\}_n, \{\delta_n\}_n) \to (\{S^n\}_n, \{\delta^n\}_n)$ defined above depends functorially on f^0.
Note: If it is painful or not worth the investment of your time that it demands, you could consider looking at the remark between Theorem 7.1 and Corollary 7.2 in Serge Lang's book.

15.6. **The two ways of computing** Ext / Tor **agree.** We go back to a not necessarily commutative ring R. Recall from Lecture 14 that

$$\mathrm{Tor}_i^R(M,-) = L_i(M\otimes_R -) : R\text{-}Mod \rightsquigarrow AbGrp,\ \ \mathrm{tor}_i^R(-,N) = L_i(-\otimes_R N) : Mod\text{-}R \rightsquigarrow AbGrp.$$

We similarly defined $\mathrm{Ext}^i_R(M,N)$ and $\mathrm{ext}^i_R(M,N)$, but we might as well define these for a general abelian category $\mathcal{A}$ with enough injectives and projectives:

Notation 15.20. Given any abelian category $\mathcal{A}$ with enough injectives and projectives, the following definitions make sense (and generalize the Ext^i_R and the ext^i_R from Lecture 14):

$$\mathrm{Ext}^i_{\mathcal{A}}(M,-) = R^i(\mathrm{Hom}_{\mathcal{A}}(M,-)) : \mathcal{A} \rightsquigarrow AbGrp,\ \ \mathrm{ext}^i_{\mathcal{A}}(-,N) = R^i(\mathrm{Hom}_{\mathcal{A}}(-,N)) : \mathcal{A}^{op} \rightsquigarrow AbGrp.$$

Remark 15.21. This is the notation that is assumed in the following theorems and proofs. However, when R is commutative, the functors $\mathrm{Ext}^i_R(M,-)$, $\mathrm{ext}^i_R(-,N)$, $\mathrm{Tor}_i^R(M,-)$ and $\mathrm{tor}_i^R(-,N)$ can be viewed as functors $R\text{-}Mod \rightsquigarrow R\text{-}Mod$; it will be implicitly left to the reader to check that the proofs all adapt to this convention.

As an application of Grothendieck's theorem that effaceable/coeffaceable δ-functors are universal, we can formally state and prove the theorem from Lecture 14 (see Theorem 14.18) which said that we have identifications $\mathrm{Ext}^i_R(M,N) \cong \mathrm{ext}^i_R(M,N)$ and $\mathrm{Tor}_i^R(M,N) \cong \mathrm{tor}_i^R(M,N)$,functorially in M and N. For this, we need to first clarify the functoriality:

Notation 15.22. (i) While $\mathrm{Ext}^i_{\mathcal{A}}(M,N)$ is by design functorial in N, we realize it as functorial in M as well: given a morphism $M \to M'$ in $\mathcal{A}^{op}$, i.e., $M' \to M$ in $\mathcal{A}$, we have a natural transformation $\mathrm{Hom}_{\mathcal{A}}(M,-) \to \mathrm{Hom}_{\mathcal{A}}(M',-)$ defined by pullback along $M' \to M$, and hence by Corollary 15.16, a natural transformation $\mathrm{Ext}^i_{\mathcal{A}}(M,-) \to \mathrm{Ext}^i_{\mathcal{A}}(M',-)$ for each i. Now it is easy to see that each $\mathrm{Ext}^i_{\mathcal{A}}$ is a functor

$$\mathcal{A}^{op} \times \mathcal{A} \rightsquigarrow AbGrp.$$

(ii) Similarly, we extend our definitions of $\mathrm{ext}^i_{\mathcal{A}}, \mathrm{Tor}_i^R$ and tor_i^R to be functors:

$$\mathcal{A}^{op} \times \mathcal{A} \rightsquigarrow AbGrp, \quad Mod\text{-}R \times R\text{-}Mod \rightsquigarrow AbGrp, \quad \text{and} \quad Mod\text{-}R \times R\text{-}Mod \rightsquigarrow AbGrp.$$

Theorem 15.23. *(i) Let $\mathcal{A}$ be an abelian category with enough injectives and projectives. Then the functors $\mathcal{A}^{op} \times \mathcal{A} \rightsquigarrow AbGrp$ given by $\mathrm{Ext}^i_{\mathcal{A}}$ and $\mathrm{ext}^i_{\mathcal{A}}$ are naturally isomorphic.*

(ii) The functors Mod-$R \times$ R-Mod $\rightsquigarrow$ AbGrp given by Tor_i^R *and* tor_i^R *are naturally isomorphic.*

Further, when R is commutative, the resulting natural isomorphisms $\mathrm{Ext}^i_R \to \mathrm{ext}^i_R$ *and* $\mathrm{Tor}_i^R \to \mathrm{tor}_i^R$ *get enhanced to natural isomorphisms of R-Mod-valued functors.*

Notation 15.24. Once we prove the theorem, we will denote both Ext and ext by Ext, and we will denote both Tor and tor by Tor.

Exercise 15.25. Before we prove the theorem, here is an application.

(i) Show that the following are equivalent for each left R-module N_3:
  (a) N_3 is $M \otimes_R -$-acyclic, for each right R-module M.
  (b) Whenever $0 \to N_1 \to N_2 \to N_3 \to 0$ is an exact sequence of left R-modules (thus N_1 and N_2 are varying though N_3 is fixed) and M is a right R-module,
$$0 \to M \otimes_R N_1 \to M \otimes_R N_2 \to M \otimes_R N_3 \to 0$$
  is exact.
  (c) N_3 is flat.

  Hint:
  - By definition, (a) is equivalent to $\mathrm{Tor}_i^R(M, N_3)$ vanishing for each right R-module M and each $i \geqslant 1$.
  - Using the long exact sequence for $\mathrm{Tor}_i^R(M, -)$, show that (b) is equivalent to $\mathrm{Tor}_1^R(M, N_3)$ vanishing for each right R-module M.
  - Using the long exact sequence for $\mathrm{tor}_i^R(-, N_3)$, show that the condition (c) is equivalent to $\mathrm{tor}_1^R(M, N_3)$ vanishing for each right R-module M.

  Now use the theorem: since tor = Tor, (b) and (c) are equivalent. To see the equivalence of (a) and (b), use dimension shifting (which works because we have a "for all R-modules M" within these conditions). Namely, what you need to show is that if $\mathrm{Tor}_1^R(M, N_3)$ vanishes for each right R-module M, then $\mathrm{Tor}_i^R(M, N_3)$ vanishes for each right R-module M and each $i \geqslant 1$. Choose a surjection $F \to M$ with F a free right R-module, and let L be its kernel, so that we have an exact sequence $0 \to L \to F \to M \to 0$. Then the exactness of $0 = \mathrm{Tor}_2^R(F, N_3) = \mathrm{Tor}_2^R(M, N_3) \to \mathrm{Tor}_1^R(L, N_3) \to \mathrm{Tor}_1^R(F, N_3) = 0$ gives $\mathrm{Tor}_2^R(M, N_3) \cong \mathrm{Tor}_1^R(L, N_3) = 0$.

(ii) Similarly, we have an 'injective version': show that in any abelian category $\mathcal{A}$, an object I is injective if and only if it is $\mathrm{Hom}_{\mathcal{A}}(A, -)$-acyclic, for each $A \in \mathrm{Ob}\,\mathcal{A}$.

(iii) Similarly, we have a 'projective version': show that in any abelian category $\mathcal{A}$, an object P is projective if and only if it is $\mathrm{Hom}_{\mathcal{A}}(-, A)$-acyclic, for each $A \in \mathrm{Ob}\,\mathcal{A}$.

The first of these exercises shows how Tor lets one deduce the exactness of certain sequences with just 'pure thought', avoiding what might have been more complicated diagram chases.

Proof of Theorem 15.23. We will prove (i); the proof of (ii) is analogous. We will only prove the *AbGrp*-valued version dealing with a general abelian category $\mathcal{A}$; the proof of the *R-Mod*-valued version dealing with *R-Mod* for commutative R is analogous. The

following proof is ultimately unsatisfactory, and also non-standard, but it is kind of 'soft' and non-irritating (I think), despite its seeming length.

For fixed $M \in \operatorname{Ob}\mathcal{A} = \operatorname{Ob}\mathcal{A}^{op}$ and varying i, we consider $\operatorname{ext}^i_{\mathcal{A}}(M,-) : N \rightsquigarrow \operatorname{ext}^i_{\mathcal{A}}(M,N)$. Let us prove that these extend to an effaceable δ-functor $(\{\operatorname{ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i)$.

We have identifications

$$\operatorname{ext}^i_{\mathcal{A}}(M,N) = H^i(\operatorname{Hom}_{\mathcal{A}}(P_\bullet, N)) = \frac{\ker(\operatorname{Hom}_{\mathcal{A}}(P_i,N) \to \operatorname{Hom}_{\mathcal{A}}(P_{i+1},N))}{\operatorname{im}(\operatorname{Hom}_{\mathcal{A}}(P_{i-1},N) \to \operatorname{Hom}_{\mathcal{A}}(P_i,N))}, \tag{72}$$

where $P_\bullet \to M$ is a projective resolution of M.

Moreover, this identification is functorial in N, by Remark 15.17. Let us write out concretely what this means: given $N \to N'$, the identifications of (72) for N and N' transport the map $\operatorname{ext}^i_{\mathcal{A}}(M,N) \to \operatorname{ext}^i_{\mathcal{A}}(M,N')$ given by Corollary 15.16, to the map $H^i(\operatorname{Hom}_{\mathcal{A}}(P_\bullet,N)) \to H^i(\operatorname{Hom}_{\mathcal{A}}(P_\bullet,N'))$ obtained by applying H^i to the 'compose with $N \to N'$ morphism' of complexes $\operatorname{Hom}_{\mathcal{A}}(P_\bullet,N) \to \operatorname{Hom}_{\mathcal{A}}(P_\bullet,N')$.

Given an exact sequence $0 \to N_1 \to N_2 \to N_3 \to 0$ in $\mathcal{A}$, we get a sequence of complexes

$$0 \to \operatorname{Hom}_{\mathcal{A}}(P_\bullet,N_1) \to \operatorname{Hom}_{\mathcal{A}}(P_\bullet,N_2) \to \operatorname{Hom}_{\mathcal{A}}(P_\bullet,N_3) \to 0. \tag{73}$$

In fact, this sequence is exact: it is termwise exact because each $\operatorname{Hom}_{\mathcal{A}}(P_i,-)$ is exact, P_i being projective.

The long exact sequence for cohomology associated to this short exact sequence then gives a long exact sequence:

$$0 \to \operatorname{Hom}_{\mathcal{A}}(M,N_1) \to \ldots\ldots\cdots \to \operatorname{ext}^i_{\mathcal{A}}(M,N_1) \to \operatorname{ext}^i_{\mathcal{A}}(M,N_2) \to \operatorname{ext}^i_{\mathcal{A}}(M,N_3) \xrightarrow{\delta^i} \operatorname{ext}^{i+1}_{\mathcal{A}}(M,N_1) \to \ldots.$$

Moreover, this long exact sequence is functorial in the short exact sequence $0 \to N_1 \to N_2 \to N_3 \to 0$, because the sequence of complexes (73) is.

Thus, we have proved that for fixed M, the $\operatorname{ext}^i_{\mathcal{A}}(M,-)$ extend to a δ-functor $(\{\operatorname{ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i)$. Moreover, this δ-functor is effaceable, because if $N = I \in \operatorname{Ob}\mathcal{A}$ is injective, then, $\operatorname{Hom}_{\mathcal{A}}(-,I)$ being exact, $\operatorname{Hom}_{\mathcal{A}}(P_{i-1},I) \to \operatorname{Hom}_{\mathcal{A}}(P_i,I) \to \operatorname{Hom}_{\mathcal{A}}(P_{i+1},I)$ is exact for $i \geqslant 1$. Therefore, by Theorem 15.18, this δ-functor is universal. Thus, we now have two universal δ-functors $(\{\operatorname{Ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i)$ and $(\{\operatorname{ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i)$. Moreover, they both define the same functor for $i = 0$, namely, $\operatorname{Hom}_{\mathcal{A}}(M,-)$. Thus, by their universality, there exists an isomorphism between these δ-functors, and in particular, natural isomorphisms $\operatorname{Ext}^i_{\mathcal{A}}(M,-) \to \operatorname{ext}^i_{\mathcal{A}}(M,-)$ for each i. All this holds for each fixed $M \in \operatorname{Ob}\mathcal{A}$.

We now need to show that these isomorphisms are functorial in M. So let $M' \to M$ be a morphism in $\mathcal{A}$. Consider the following diagram of δ-functors $\mathcal{A} \to AbGrp$:

$$\begin{array}{ccc} (\{\operatorname{Ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i) & \longrightarrow & (\{\operatorname{ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i) \\ \downarrow & & \downarrow \\ (\{\operatorname{Ext}^i_{\mathcal{A}}(M',-)\}_i, \{\delta^i\}_i) & \longrightarrow & (\{\operatorname{ext}^i_{\mathcal{A}}(M',-)\}_i, \{\delta^i\}_i) \end{array},$$

where each arrow is obtained from the universality of its source, corresponding to either $\mathrm{Hom}_{\mathcal{A}}(M,-) \to \mathrm{Hom}_{\mathcal{A}}(M,-)$ or $\mathrm{Hom}_{\mathcal{A}}(M,-) \to \mathrm{Hom}_{\mathcal{A}}(M',-)$ in degree 0. The horizontal arrows are then formed of the natural isomorphisms $\mathrm{Ext}^i_{\mathcal{A}}(M,-) \to \mathrm{ext}^i_{\mathcal{A}}(M,-)$ and $\mathrm{Ext}^i_{\mathcal{A}}(M',-) \to \mathrm{ext}^i(\mathcal{A},-)$ described above. Note that the left vertical arrow is obtained by applying Corollary 15.16 to the natural transformation $\mathrm{Hom}_{\mathcal{A}}(M,-) \to \mathrm{Hom}_{\mathcal{A}}(M',-)$. The right vertical arrow is readily verified to be given by taking the cohomology of a morphism $P'_\bullet \to P_\bullet$ of projective resolutions lifting $M' \to M$: in other words, the isomorphisms $\mathrm{ext}^i_{\mathcal{A}}(M,N) \to \mathrm{ext}^{\mathcal{A}}_i(M',N)$ given by it are the same as the ones obtained from the functoriality of $\mathrm{ext}^{\mathcal{A}}_i(-,N)$.

Therefore, it is now enough to prove the commutativity of the above diagram. But this follows from the fact that the compositions $(\{\mathrm{Ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i) \to (\{\mathrm{ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i) \to (\{\mathrm{ext}^i_{\mathcal{A}}(M',-)\}_i, \{\delta^i\}_i)$ and $(\{\mathrm{Ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i) \to (\{\mathrm{Ext}^i_{\mathcal{A}}(M',-)\}_i, \{\delta^i\}_i) \to (\{\mathrm{ext}^i_{\mathcal{A}}(M',-)\}_i, \{\delta^i\}_i)$ are both morphisms of δ-functors, with the source $(\{\mathrm{Ext}^i_{\mathcal{A}}(M,-)\}_i, \{\delta^i\}_i)$ universal, and the composites are the same in degree 0, namely, $\mathrm{Hom}_{\mathcal{A}}(M,-) \to \mathrm{Hom}_{\mathcal{A}}(M',-)$. □

Remark 15.26. (i) For $i = 0$, the isomorphisms yielded by the above proof are just the identity maps $\mathrm{Hom}_{\mathcal{A}}(M,N) \to \mathrm{Hom}_{\mathcal{A}}(M,N)$.

(ii) The above proof yields something more precise than the statement of the theorem: it extends the $\mathrm{ext}^i_{\mathcal{A}}(M,N)$, i varying over $\mathbb{Z}$, into a cohomological δ-functor in a particular way, and then gives a *unique* collection of functorial isomorphisms $\mathrm{Ext}^i_{\mathcal{A}}(M,N) \to \mathrm{ext}^i_{\mathcal{A}}(M,N)$ subject to two constraints: the requirement of compatibility with the δ-maps, and the requirement of being, in degree 0, the identity maps $\mathrm{Hom}_{\mathcal{A}}(M,N) \to \mathrm{Hom}_{\mathcal{A}}(M,N)$.

15.7. **Applications to** Ext **and** Tor**.** The following can be proved as an application, but it is also easy to see it directly (as described below):

Proposition 15.27. *(i) Let $\mathcal{A}$ be an abelian category. For all collections $\{M_i\}_i$ of objects of $\mathcal{A}$, and all $N \in \mathrm{Ob}\,\mathcal{A}$, we have isomorphisms functorial in the M_i and N:*

$$\mathrm{Ext}^i_{\mathcal{A}}\Big(\bigoplus_i M_i, N\Big) \cong \prod_i \mathrm{Ext}^i_{\mathcal{A}}\Big(M_i, N\Big),$$

$$\mathrm{Ext}^i_{\mathcal{A}}\Big(N, \prod_i M_i\Big) \cong \prod_i \mathrm{Ext}^i_{\mathcal{A}}\Big(M_i, N\Big).$$

(ii) For all collections $\{M_i\}_i$ of right R-modules and all left R-modules N, we have isomorphisms functorial in the M_i and N:

$$\mathrm{Tor}^R_i\Big(\bigoplus_i M_i, N\Big) \cong \bigoplus_i \mathrm{Tor}^R_i\Big(M, N\Big),$$

and an analogous assertion holds with respect to direct sums in the second factor.

Proof. Easy: Use that the homology/cohomology of a direct sum/product of complexes is the direct sum/product of homologies/cohomologies. Note the source of the apparent asymmetry here: $\mathrm{Hom}_{\mathcal{A}}(\bigoplus_i M_i, N) \cong \prod_i \mathrm{Hom}_{\mathcal{A}}(M_i, N)$ and $\mathrm{Hom}_{\mathcal{A}}(N, \prod_i M_i) \cong \prod_i \mathrm{Hom}_{\mathcal{A}}(N, M_i)$. □

Proposition 15.28. *Assume that $R \to S$ is a homomorphism of commutative rings, such that S/R is flat. Then:*

(i) For all $i \in \mathbb{N}$, we have, functorially in R-modules M and N:

$$\mathrm{Tor}_i^R(M \otimes_R S, N \otimes_R S) \cong \mathrm{Tor}_i^R(M, N) \otimes_R S.$$

(ii) Assume that R is Noetherian. Then we have, for all $i \in \mathbb{N}$, functorially in R-modules N and finitely generated R-modules M:

$$\mathrm{Ext}^i_R(M, N) \otimes_R S \cong \mathrm{Ext}^i_S(M \otimes_R S, N \otimes_R S).$$

The hard work done earlier regarding the universality of effaceable and coeffaceable δ-functors reduces (modulo some standard checking) the proofs of various propositions (such as the above one) to checking them in degree 0. In particular, when one sees a proposition such as the above, it would be natural to start out by trying out the case where $i = 0$, which involves functors that we are familiar with. For Tor, this is one of the easy properties of the tensor product, while for Ext, this is the following lemma.

Lemma 15.29. *Let $R \to S$ be a homomorphism of commutative rings, such that S/R is flat. Then we have isomorphisms, funtorially in R-modules N and finitely presented R-modules M:*

$$\mathrm{Hom}_R(M, N) \otimes_R S \cong \mathrm{Hom}_S(M \otimes_R S, N \otimes_R S).$$

Proof. The assertion is easy for finitely generated free R-modules M: if $M \cong R^n$, then both sides have an obvious isomorphism with $(N \otimes_R S)^n$, and the functoriality in morphisms of finitely generated free R-modules is easy to check. In general, to say that M is finitely presented means that $M = \mathrm{coker}(F_1 \to F_0)$, for some finitely generated free R-modules F_1 and F_0. The free case that was just discussed gives the non-dotted vertical arrows in the following commutative diagram:

$$\begin{array}{ccccccc}
0 & \longrightarrow & \mathrm{Hom}_R(M, N) \otimes_R S & \longrightarrow & \mathrm{Hom}_R(F_0, N) \otimes_R S & \longrightarrow & \mathrm{Hom}_R(F_1, N) \otimes_R S \\
 & & \vdots & & \downarrow \cong & & \downarrow \cong \\
0 & \longrightarrow & \mathrm{Hom}_S(M \otimes_R S, N \otimes_R S) & \longrightarrow & \mathrm{Hom}_S(F_0 \otimes_R S, N \otimes_R S) & \longrightarrow & \mathrm{Hom}_R(F_1 \otimes_R S, N \otimes_R S)
\end{array}.$$

The above diagram has exact rows, since $\mathrm{Hom}_R(-, N)$ and $\mathrm{Hom}_S(-, N \otimes_R S)$ are left exact functors, while $- \otimes_R S$ is exact (because S is assumed to be flat over R). Thus, since the second and the third vertical arrows are isomorphisms, an isomorphism is induced between $\mathrm{Hom}_R(M, N) \otimes_R S$ and $\mathrm{Hom}_S(M \otimes_R S, N \otimes_R S)$, as desired. □

Proof of Proposition 15.28. Let us prove the assertion about Ext; the proof of the assertion about Tor is similar, but easier. Fixing N, we view both sides as functors from the category $((R\text{-}Mod)^{fg})^{op}$, which is opposite to the category of finitely generated R-modules, to the category of S-modules. Since R is Noetherian, $((R\text{-}Mod)^{fg})^{op}$ is an abelian category ($(R\text{-}Mod)^{fg}$ being a full additive subcategory of $R\text{-}Mod$ closed under kernels and cokernels). Clearly, this category has enough projectives.

Each side represents a collection of functors $((R\text{-}Mod)^{fg})^{op} \rightsquigarrow S\text{-}Mod$, indexed by varying i, that fits into a δ-functor: for the left-hand side, this uses that tensoring with the flat R-algebra S sends long exact sequences of R-modules to those of S-modules, while for the right-hand side, this uses that tensoring with the flat R-algebra S sends short exact sequences of R-modules to those of S-modules. Moreover, each of these δ-functors, in degrees $\geqslant 1$, vanishes on injective objects of $((R\text{-}Mod)^{fg})^{op}$, i.e., on projective R-modules (for the functors on the right-hand side, note and use that $- \otimes_R S$ sends projective R-modules to projective S-modules), so they are effaceable. Hence they are universal by Grothendieck's theorem (Theorem 15.18), so to show that they are naturally isomorphic, it is enough to show that the two sides are naturally isomorphic in degree 0. This is taken care of by Lemma 15.29.

In the Tor case, the proof is similar, but easier: one does not resrict to finitely generated R-modules, and in degree 0 one uses

$$(M \otimes_R N) \otimes_R S \cong (M \otimes_R N) \otimes_S (S \otimes_S S) \cong (M \otimes_R S) \otimes_S (N \otimes_R S).$$

□

Proposition 15.30. *For all $i \geqslant 0$, we have isomorphisms functorial in right R-modules M and left R-modules N:*

$$\mathrm{Tor}_i^R(M, N) \cong \mathrm{Tor}_i^R(N, M).$$

Sketch of proof. We will show this for fixed M and varying N; the general case where M is allowed to vary can be handled by an argument similar to that seen towards the end of the proof of Theorem 15.23.

For fixed M and varying N and i, both sides are part of homological δ-functors that are coeffaceable and hence universal. Thus, we are reduced to showing that they are naturally isomorphic in degree 0, where we have the isomorphism $M \otimes_R N \cong N \otimes_R M$. □

Remark 15.31. Unlike the tensor product, Tor does not commute with arbitrary colimits. Indeed, a cokernel is a colimit, but if $\mathrm{Tor}_1^R(M_1, N) = \mathrm{Tor}_1^R(M_2, N) = 0$ (which is true if M_1 and M_2 are flat), it does not follow that $\mathrm{Tor}_1^R(\mathrm{coker}(M_1 \to M_2), N) = 0$ ($\mathrm{coker}(M_1 \to M_2)$ may not be flat). However, we saw in Proposition 15.27 that Tor does commute with direct sums, and we will see in Proposition 15.32 below that Tor commutes with directed colimits.

Proposition 15.32. Tor *commutes with directed colimits. In other words, let R be a ring, J a directed set viewed as a category as in Lecture 4 (thus, there is exactly one morphism $i \to j$ if $i \leqslant j$, and none otherwise), and consider a directed system $J \rightsquigarrow R\text{-}Mod$, written more informally as $j \rightsquigarrow M_j$. Note that the maps $\mathrm{Tor}_i^R(M_j, N) \to \mathrm{Tor}_i^R\Big(\varinjlim_l M_l, N\Big)$ (obtained by applying the functoriality of Tor_i^R in the first variable to $M_j \to \varinjlim_l M_l$) give by the definition of directed colimits a map*

$$\varinjlim_l \mathrm{Tor}_i^R(M_l, N) \to \mathrm{Tor}_i^R\Big(\varinjlim_l M_l, N\Big).$$

This map is an isomorphism.

The proof of Proposition 15.32 will use:

Lemma 15.33. *Let J be a directed set viewed as a category. Consider J-indexed directed systems in R-Mod, namely* $\mathrm{Fun}(J, R\text{-}Mod)$*, which is an abelian category. Then 'taking directed colimits' is a functor*

$$\mathrm{Fun}(J, R\text{-}Mod) \rightsquigarrow R\text{-}Mod$$

(recall that small colimits exist in R-Mod). This functor is exact.

Remark 15.34. The above result would not be true with R-Mod replaced by some other category where directed colimits exist. For instance, it is not true for $(R\text{-}Mod)^{op}$, because you can check that inverse limits in R-Mod are not exact.

Proof of Lemma 15.33. In the lecture I gave it as an exercise, here too you can see it directly, but I will give a proof parts of which are less direct.

If you have difficulty seeing that 'taking the directed colimit' is a functor, recall that if a directed system is given in the more usual notation by $(\{M_j\}_j, (\varphi_{ji} : M_i \to M_j))$, then a directed colimit can be described as

$$\Big(\bigoplus_j M_j\Big)/\mathrm{Span}_R(\{m_i - \varphi_{ji}(m_i) \mid i \leqslant j, m_i \in M_i\}).$$

From this, the functoriality is especially easy to see (though one can do this by diagram chasing, which works in an arbitrary abelian category where directed colimits exist).

Of this, the right exactness is a consequence of some generalities (and holds in a more general setting than our R-Mod one, one where just the colimits are required to exist): Let

$$0 \to (L_j)_j \to (M_j)_j \to (N_j)_j \to 0$$

be an exact sequence of directed systems indexed by J; we have suppressed the transition maps from notation for brevity. Since exactness in $\mathrm{Fun}(J, R\text{-}Mod)$ is determined pointwise, $0 \to L_j \to M_j \to N_j \to 0$ is exact for each j. For the right exactness, it is enough to show that for each R-module K,

$$\mathrm{Hom}_R\Big(\varinjlim_j N_j, K\Big) \to \mathrm{Hom}_R\Big(\varinjlim_j M_j, K\Big)$$

is injective. But by the definition of a colimit ($\mathrm{Hom}(-, K)$ converts it into a limit), this identifies with a map

$$\varprojlim_j \mathrm{Hom}_R(N_j, K) \to \varprojlim_j \mathrm{Hom}_R(M_j, K),$$

which, by the explicit description of inverse limit in Set, is manifestly injective.

For the left exactness, which is what is special to R-Mod, suppose l in $\varinjlim L_j$ maps to 0 in $\varinjlim M_j$. Then l is the image of some $l_j \in L_j$, which maps to say $m_j \in M_j$. By assumption,

m_j maps to 0 in some M_k, $k > l$. But this means that l_j maps to 0 in L_k, since $L_k \to M_k$ is injective. Thus, $l = 0$, showing that $\varinjlim L_j \to \varinjlim M_j$ is injective. □

Proof of Proposition 15.32. For a fixed directed system $j \rightsquigarrow M_j$ with directed colimit M, both sides are functors in N. The right-hand side, as i varies, clearly forms a universal δ-functor. So does the left-hand side: this follows from Lemma 15.33, which implies that the directed colimit of the directed system, in $l \in J$, of long exact sequences obtained by applying the $\mathrm{Tor}_i^R(M_l, -)$ to a short exact sequence $0 \to N_1 \to N_2 \to N_3 \to 0$ is indeed long exact. Moreover, both sides are coeffaceable functors, since both sides vanish when N is projective (or even flat).

Therefore, it suffices to show that the two sides are naturally isomorphic in degree 0, which follows from the commutativity of tensor product with colimits. □

Corollary 15.35. *Over a (commutative) PID R, every torsion-free module M is flat.*

Proof. The flatness of M is equivalent to the assertion that $\mathrm{Tor}_1^R(M, N) = 0$ for all R-modules N (that this is so is an easy consequence of the long exact sequence for Tor, and is anyway recalled in Proposition 15.36 below). If M is finitely generated, the corollary follows from the fact that finitely generated torsion-free modules over a PID are free. The same then follows for any torsion-free M, since any M is a directed colimit of its finitely generated submodules (take J to be the directed set of finitely generated submodules of M, partially ordered under inclusion, and set $M_j = j$ for each $j \in J$), and Tor commutes with directed colimits. □

Proposition 15.36. *If M is a module over a commutative ring R, the following are equivalent:*

(i) M *is flat.*
(ii) $\mathrm{Tor}_i^R(M, N) = 0$ *for all* $i \geqslant 1$ *and* R*-modules* N*.*
(iii) $\mathrm{Tor}_1^R(M, N) = 0$ *for all* R*-modules* N*.*
(iv) $\mathrm{Tor}_1^R(M, R/I) = 0$ *for all ideals* $I \subset R$*.*
(v) $\mathrm{Tor}_1^R(M, R/I) = 0$ *for all finitely generated ideals* $I \subset R$*.*

Proof. If M is flat, then by the long exact sequence for Tor associated to some exact sequence of the form $0 \to N' \to F \to N \to 0$, where F is free, so that $\mathrm{Tor}_1^R(M, F) = 0$, we get $\mathrm{Tor}_1^R(M, N) \cong \ker(M \otimes_R N' \to M \otimes_R F) = 0$. Thus, $\mathrm{Tor}_1^R(M, N) = 0$ for all N. Inductively, if $\mathrm{Tor}_i^R(M, -)$ vanishes, then given any R-module N, the same exact sequence as above gives $\mathrm{Tor}_i^R(M, N) \cong \mathrm{Tor}_{i-1}^R(M, N') = 0$. Thus, (i) implies (ii). It is trivial that (ii) implies (iii).

Suppose (iii) holds. Applying the long exact sequence for Tor to $0 \to N_1 \to N_2 \to N_3 \to 0$, and using that $\mathrm{Tor}_1^R(M, N_3) = 0$, we get that $0 \to M \otimes_R N_1 \to M \otimes_R N_2 \to M \otimes_R N_3 \to 0$ is exact. Thus, (iii) implies (i). Thus, we conclude that (i), (ii) and (iii) are equivalent.

It is trivial that (iii) implies (iv). Conversely, suppose (iv) holds. To show that $\mathrm{Tor}_1^R(M,N) = 0$ for all R-modules N, the commutativity of Tor with directed colimits (Proposition 15.32) reduces us to the case where N is finitely generated (as in the proof of Corollary 15.35).

We wish to reduce N further. Now if $0 \to N' \to N \to N'' \to 0$ is exact, the long exact sequence for Tor gives the exactness of $\mathrm{Tor}_1^R(M,N') \to \mathrm{Tor}_1^R(M,N) \to \mathrm{Tor}_1^R(M,N'')$, so the vanishing of $\mathrm{Tor}_1^R(M,N)$ follows from that for $\mathrm{Tor}_1^R(M,N')$ and $\mathrm{Tor}_1^R(M,N'')$. Since an R-module N generated by n elements fits into an exact sequence $0 \to N' \to N \to N'' \to 0$ where each of N' and N'' is generated by fewer than n elements, an iterated application of the preceding statement implies that (iii) is equivalent to the vanishing of $\mathrm{Tor}_1^R(M,N)$ for all cyclic (i.e., singly generated) R-modules N, which are precisely those of the form R/I. Thus, (iv) is equivalent to the conditions that precede it.

Now by the long exact sequence for Tor, this time applied to $0 \to I \to R \to R/I \to 0$, we have the formula $\mathrm{Tor}_1^R(M,R/I) \cong \ker(I \otimes_R M \to M)$. By the commutativity of tensor products with colimits, $I \otimes_R M$ is simply the directed colimit of the $I_1 \otimes_R M$ as I_1 ranges over the finitely generated ideals contained in I. By the exactness of the "take the directed colimit" functor (in the context of R-Mod; see Lemma 15.33), if each of these $I_1 \otimes_R M \to M$ is injective, so is $I \otimes_R M \to M$. Therefore, we conclude that (v) is equivalent to (iv). $\square$

Exercise 15.37. Rephrase Baer's criterion for the injectivity of an R-module M as follows: M is injective if and only if $\mathrm{Ext}_R^1(R/I,M) = 0$ for all ideals $I \subset R$.

16. Lecture 16 — Composition series, semisimplicity, Jacobson radical

16.1. **Composition series and the Jordan-Hölder theorem.** Today, unless otherwise stated, R will denote a ring that is not assumed to be commutative. We will discuss various results for left R-modules, where R is a ring. It will be implicitly assumed that analogous results apply to right R-modules.

Many of the following notions (semisimple, Artinian, Noetherian, composition series etc.) will also apply in the context of an abelian category. For simplicity we will only state them for modules; when needed you can work out/look up other abelian category versions.

Definition 16.1. A left R-module M is called simple, or in some situations irreducible, if it is nonzero and has no proper nonzero submodules.

Exercise 16.2. (i) Show that each simple left R-module is of the form $R/\mathfrak{m}$ for some maximal left ideal $\mathfrak{m} \subset R$, and conversely, that for each maximal left ideal $\mathfrak{m} \subset R$, the left R-module $R/\mathfrak{m}$ is simple.

(ii) However, it is not the case that isomorphism classes of simple left R-modules are in bijection with maximal left ideals $\mathfrak{m} \subset R$: this is because for distinct maximal left ideals $\mathfrak{m}_1, \mathfrak{m}_2 \subset R$, the left R-modules $R/\mathfrak{m}_1$ and $R/\mathfrak{m}_2$ can be isomorphic, for instance if $\mathfrak{m}_1 = \mathfrak{m}_2 a$ for some left and right invertible $a \in R$. [48]
Show, nevertheless: there is a bijection between maximal left ideals $\mathfrak{m} \subset R$ and isomorphism classes of pairs (M, m), where M is a simple left R-module, and $0 \neq m \in M$ (it is your task to make precise what these isomorphism classes mean).
Hint: Send (M, m) to $\mathrm{Ann}_R(m) \subset R$, a left ideal. This is motivated by Lemma 16.29 below.

Please keep in mind that if R is not commutative, $R/\mathfrak{m}$ will not be a ring, and hence not a field or anything. For instance, see the first example below.

Example 16.3. (i) If $R = M_n(k)$ with k a field, show as an easy exercise that k^n, viewed as $1 \times n$ column vectors with the standard (matrix multiplication) action of $R = M_n(k)$, is a simple left R-module. We will hopefully see later that this is only only simple left R-module up to isomorphism.

(ii) A left ideal $I \subset R$ is a simple R-module if and only if it is a minimal left ideal of R (by which, we mean minimal among the nonzero left ideals of R).

(iii) If k is a field and G is a finite group, then a representation V of G is irreducible if and only if, viewed as a $k[G]$-module, V is simple.

Remark 16.4. For those who know the Nullstellensatz: the set of isomorphism classes of simple modules is thus a noncommutative version of the 'spectrum' of a variety.

[48] This problem does not arise when R is commutative: when R is commutative, $R/\mathfrak{m}$ determines $\mathrm{Ann}_R(R/\mathfrak{m}) = \mathfrak{m}$, but for more general R, $\mathrm{Ann}_R(R/\mathfrak{m})$, being a two-sided ideal, is typically smaller than $\mathfrak{m}$. We will use this later in this lecture.

Definition 16.5. (i) A composition series for a left R-module M is a sequence of submodules $(M_i)_{1\leqslant i\leqslant r}$ of M, for which there are inclusions

$$0 = M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots \subsetneq M_r = M,$$

such that each M_i/M_{i-1} $(1 \leqslant i \leqslant r)$ is a simple left R-module. We call r the length of this series.

(ii) Two composition series $(M_i)_{1\leqslant i\leqslant r}$ and $(M'_i)_{1\leqslant i\leqslant s}$ for M are said to be equivalent if we have an equality of *multisets* [49]

$$\{[M_1/M_0], \ldots, [M_r/M_{r-1}]\} = \{[M'_1/M'_0], \ldots, [M'_s/M'_{s-1}]\},$$

where for each R-module N, we write $[N]$ for the isomorphism class of N.

Equivalently, if $r = s$, and there exists a permutation $\sigma \in S_r$ such that for each $1 \leqslant i \leqslant r$, $M_i/M_{i-1} \cong M'_{\sigma(i)}/M'_{\sigma(i)-1}$.

(iii) A left R-module is said to have finite length if it has a composition series.

Example 16.6. The two composition series of the $\mathbb{Z}$-module $\mathbb{Z}/6\mathbb{Z}$, given by

$$0 \subset 3\cdot\mathbb{Z}/6\mathbb{Z} \subset \mathbb{Z}/6\mathbb{Z}, \qquad \text{and} \qquad 0 \subset 2\cdot\mathbb{Z}/6\mathbb{Z} \subset \mathbb{Z}/6\mathbb{Z},$$

are equivalent: the successive quotients for the first series are $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ (in that order), while those for the second series are $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$.

Theorem 16.7 (Jordan-Hölder)**.** *If a left R-module M has finite length, then all composition series for M are equivalent, and in particular have the same length.*

Proof. Since this is standard, I am being slightly sloppy. Suppose $(M_i)_{i=1}^r$ and $(M'_i)_{i=1}^s$ are two composition series for M. We will use induction on r. If $r = 0$ (resp., $r = 1$), then M is zero (resp., M is simple), and the theorem is trivial/easy.

Let $1 \leqslant j \leqslant r$ be the smallest such that $M'_1 \subset M_j$. Then M/M'_1 has a composition series

$$0 \subset (M_1+M'_1)/M'_1 \subsetneq \cdots \subsetneq (M_{j-2}+M'_1)/M'_1 \subsetneq (M_{j-1}+M'_1)/M'_1 = M_j/M'_1 \subsetneq \cdots \subsetneq M_r/M'_1$$

(if $j = 1$, the series begins at $0 = (M_0 + M'_1)/M'_1 = M_1/M_1$), which has length $r - 1$, as well as

$$0 \subsetneq M'_2/M'_1 \subsetneq \cdots \subsetneq M'_s/M'_1,$$

which has length $s - 1$. It is clear that the multisets of successive quotients of each of the given composition series for M are obtained by taking the union of $[M'_1] = [M_j/M_{j-1}]$ and the corresponding composition series for M/M'_1. Thus, by induction the theorem follows. □

Remark 16.8. (i) The proof shows that if $(M_i)_{i=1}^r$ is a composition series for M, then given any sequence $0 = M'_0 \subsetneq M'_1 \subsetneq M'_2 \subsetneq \cdots \subsetneq M'_s = M$ of submodules of M, such that M'_i/M'_{i-1} is simple for all $1 \leqslant i \leqslant s$, we have $s \leqslant r$, and that $\{[M'_i/M'_{i-1} \mid 1 \leqslant i \leqslant s\}$ is a sub-multiset of $\{[M_i/M_{i-1}] \mid 1 \leqslant i \leqslant r\}$. In other

[49] Multisets are "sets where multiplicity is allowed"; since we will only be bothered with finite multisets, a formal way to describe a multiset of the kind we will care about is as a set X together with a 'multiplicity function' $X \to \mathbb{Z}_{\geqslant 0}$.

words, one M has a composition series, it has no "infinite length analogue" of a composition series.

(ii) There is no naive generalization of Theorem 16.7 to infinite length M: The $\mathbb{Z}$-module $\mathbb{Z}$, which is not of finite length, has "infinite analogues of composition series"

$$\cdots \subsetneq q^2\mathbb{Z} \subsetneq q\mathbb{Z} \subsetneq \mathbb{Z}, \qquad \ldots p^2\mathbb{Z} \subsetneq p\mathbb{Z} \subsetneq \mathbb{Z},$$

for any two primes p and q; the simple subquotients associated to these series are all isomorphic to $\mathbb{Z}/q\mathbb{Z}$ for the first series and $\mathbb{Z}/p\mathbb{Z}$ for the second, and thus the series are not equivalent in any reasonable sense if $p \neq q$.

Rishiraj was not happy with the above example since it did not 'begin anywhere', despite 'decreasing to 0'. This is a valid point, and from quick thought I don't seem to have counterexamples that begin at something simple and proceed exhaustingly.

Definition 16.9. If a left R-module M has finite length, we choose a composition series $0 = M_0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_r = M$ for M, and define:

(i) $JH(M)$, the multiset of composition factors or Jordan-Hölder factors or Jordan-Hölder constituents of M, to be the multiset $\{[M_i/M_{i-1}] \mid 1 \leqslant i \leqslant r\}$

(ii) $l(M)$, the length of M, to be r.

Then $JH(M)$ and $l(M)$ are independent of the choices of the composition series $0 = M_0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_r = M$, by Theorem 16.7. If M does not have a composition series, we set $l(M) = \infty$.

The following lemma is immediate from the isomorphism theorems:

Lemma 16.10. *If M is an R-module and $N \subset M$ is a submodule, then $l(M) < \infty$ if and only if the two conditions $l(N) < \infty$ and $l(M/N) < \infty$ are satisfied. Morever, when these conditions are satisfied, we have $JH(M) = JH(N) \cup JH(M/N)$ (union as multisets), and $l(M) = l(N) + l(M/N)$.*

Proof. Easy. □

Example 16.11. (i) $l(0) = 0$, and the left R-modules of length 1 are precisely the simple left R-modules.

(ii) If $R = k$ is a field, and V is a vector space over k, then $l(V) < \infty$ if and only if $\dim_k V < \infty$, in which case $l(V) = \dim V$.

(iii) If $R = \mathbb{Z}$, show as an easy exercise that $l(M) < \infty$ if and only if M is a finite abelian group. In particular $l(\mathbb{Z}) = \infty$, which is anyway forced by Remark 16.8(ii).

(iv) Even when M has finite length, $JH(M)$ does not determine M: with $R = \mathbb{Z}$, we have $JH(\mathbb{Z}/4\mathbb{Z}) = JH(\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z})$. A related observation is that M may not be a direct sum of its Jordan-Hölder factors, but it is made up from its Jordan-Hölder factors by extensions that may be nontrivial.

(v) In case M is known to be a finite direct sum of simple modules, then it is immediate that $JH(M)$ determines M. This condition on M is that of semisimplicity, which we will study later.

(vi) The proof of the Jordan-Hölder theorem, Theorem 16.7, can be adapted to the setting of finite not necessarily abelian groups G, where a composition series refers to a chain $0 = G_0 \subsetneq G_1 \subsetneq \cdots \subsetneq G_r = G$ such that each $G_{i-1} \subset G_i$ is normal, and G_i/G_{i-1} is simple (a group is said to be simple if it has no nontrivial proper normal subgroup). The version of the Jordan-Hölder theorem for this setting says that, while G might have many composition series $0 = G_0 \subsetneq G_1 \subsetneq \cdots \subsetneq G_r = G$, they are all equivalent in the sense that $\{[G_i/G_{i-1}] \mid 1 \leqslant i \leqslant r\}$ is independent of the chosen series.

Remark 16.12. Here are two applications of the notion of length:

(i) A finite dimensional representation V of a finite group G over a field k, viewed as a $k[G]$-module, clearly has finite length. It may not be a direct sum of irreducible representations, but it is built from the irreducible representations that are its Jordan-Hölder factors, as observed above. In this case, $l(V)$ is the number of irreducible representations of which G is formed of.

(ii) Intersection multiplicities (this discussion is highly informal). In $\mathbb{C}^2$, consider the intersection of $x^2 + y^2 = 1$ and $x = 0$ – there are two points of intersection, $(0, 1)$ and $(0, -1)$. On the other hand, $x^2 + y^2 = 1$ and $x = 1$ have only one intersection, $(1, 0)$. We want a degree 2 curve to intersect a degree 1 curve in two points, and it does seem appropriate to count $(1, 0)$ as having multiplicity 2 in the intersection between $x = 1$ and $x^2 + y^2 = 1$: after all, when we draw the picture in $\mathbb{R}^2$, $x = 1$ is a tangent to $x^2 + y^2 = 1$ at $(1, 0)$. One way to formalize this is to think of the intersection of $f = 0$ and $g = 0$ as defined by the ideal (f, g) generated by f and g: [50] in this case, taking $R = \mathbb{C}[x, y]$, we have

$$l_R(\mathbb{C}[x, y]/(x^2 + y^2 - 1, x)) = 2 = l_R(\mathbb{C}[x, y]/(x^2 + y^2 - 1, x - 1))$$

(exercise), where l_R stands for 'length as an R-module'. This modification is not enough when there are 'intersections at ∞', but that can be solved by projectivization, and generalizes to higher degree curves in the form of Bezout's theorem.

In short, the notion of length is important in describing intersection multiplicities.

More blah blah: By the right exactness of tensor product, we have $\mathbb{C}[x, y]/(f, g) = \mathbb{C}[x, y]/(f) \otimes_{\mathbb{C}[x,y]} \mathbb{C}[x, y]/(g)$. This corresponds to the fact, in the *opposite* category of affine varieties, that the variety $\{f = g = 0\}$ is the fiber product, which in this case is the intersection, of $\{f = 0\}$ and $\{g = 0\}$ over $\mathbb{C}^2$.

Please keep in mind that the above discussion is crude, and I have not gotten into the precise definitions of intersection multiplicities, which involve appropriate localizations. In higher dimensions, there is another problem associated to the tensor product behaving badly, something which Serre addressed by adding in 'Tor' terms that correct for the bad behavior of the tensor product.

We conclude with a definition that one encounters quite often:

[50] Thus, we are keeping track of the equatiohns themselves, and not just the solutions.

Definition 16.13. A subquotient of a left R-module M is a quotient module of a submodule of M.

Exercise 16.14. If M has finite length, then the simple subquotients of M are precisely the Jordan-Hölder constituents of M (up to isomorphism).

16.2. **Noetherian and Artinian rings and modules.**

Definition 16.15. (i) A left R-module M is called Artinian if it satisfies the following equivalent conditions, whose equivalence is left as an exercise:
- Every descending chain of submodules $M_1 \supseteq M_2 \supseteq \dots$ of M stabilizes, i.e., there exists $r \in \mathbb{N}$ such that $M_r = M_{r+1} = \dots$; this is often referred to as the descending chain condition (dcc).
- Any collection of submodules of M has a minimal element.

(To show that the first condition implies the second, given a collection Ξ, choose M_1 in it, if it is not minimal, then $M_2 \subsetneq M_1$ in it, and so on: to justify the 'so on', as per the comment of Professor Nitin Nitsure I told you about a few lectures ago, one needs the axiom of choice).

(ii) R is called left Artinian if R is Artinian as a left module over itself (by our convention, we similarly have right Artinian rings etc.).

(iii) Similarly, we define what it means for a left R-module M to be Noetherian, by imposing an ascending chain condition (acc).

(iv) Similarly, we define what it means for a ring R to be left Noetherian.

When R is commutative, there is no difference between 'left' and 'right', and hence the 'left' or 'right' will be dropped. In this case, the Noetherian notions were already defined in Lecture 1.

Example 16.16. Show the following as easy exercises.

(i) $\mathbb{Z}$ as a $\mathbb{Z}$-module is Noetherian, but not Artinian ($\mathbb{Z} \supsetneq p\mathbb{Z} \supsetneq p^2\mathbb{Z} \supsetneq \dots$).

(ii) If V is a vector space over a field $R = k$, then V is Noetherian if and only if it is Artinian if and only if it is finite dimensional.

(iii) $\mathbb{Z}[t]$ is Noetherian as a ring (this is a consequence of what is known as the Hilbert basis theorem, something we will hopefully see later), but it is neither Artinian nor Noetherian as a $\mathbb{Z}$-module. Similarly, $\mathbb{Q}/\mathbb{Z}$ is neither Artinian nor Noetherian as a $\mathbb{Z}$-module.

(iv) Show that $(\mathbb{Q}/\mathbb{Z})[p^\infty] = \{a/p^n \mid a, n \in \mathbb{Z}\}/\mathbb{Z}$ is Artinian as a $\mathbb{Z}$-module, but not Noetherian.

(v) Show that the ring

$$R := \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a, b \in \mathbb{R}, c \in \mathbb{Q} \right\} \subset M_2(\mathbb{R})$$

is left Artinian and left Noetherian, but neither right Artinian or right Noetherian. **Note:** For this, you can use Lemma 16.17 below. Apply it to the exact sequence

of R-modules:

$$0 \to \begin{pmatrix} 0 & * \\ 0 & 0 \end{pmatrix} \cap R \to R \to \mathbb{R} \times \mathbb{Q} \to 0,$$

where the map with target $\mathbb{R} \times \mathbb{Q}$ sends $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ to (a, c).

Now $\mathbb{R} \times \mathbb{Q}$ is both Noetherian and Artinian as an $\mathbb{R} \times \mathbb{Q}$-module, so use Lemma 16.17 to reduce to verifying that $\mathbb{R}$ is finite dimensional as an $\mathbb{R}$-vector space but infinite dimensional as a $\mathbb{Q}$-vector space.

Lemma 16.17. *Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence of left R-modules. Then:*

(i) M is Artinian if and only if M' and M'' are (everything as left R-modules).
(ii) M is Noetherian if and only if M' and M'' are.

Proof. Easy exercise. For the implications "$\Leftarrow$", given a chain in M, consider the chain in M' obtained by intersecting with M', and the chain in M'' obtained by projecting to M'': once both chains stabilize, show that the original stabilizes. □

As in the commutative Noetherian case we can use this to prove:

Lemma 16.18. *If R is left-Noetherian, then a left R-module M is Noetherian if and only if it is finitely generated.*

Proof. The ascending chain condition easily gives that any Noetherian left R-module is finitely generated. The converse is easy using Lemma 16.17. □

Lemma 16.19. *A left R-module M is of finite length if and only if it is both Artinian and Noetherian.*

Proof. "$\Rightarrow$": It is immediate that any simple left R-module is both Artinian and Noetherian. Inductive applications of Lemma 16.17 then show that any finite length left R-module is both Artinian and Noetherian.

"$\Leftarrow$": Suppose M is both Artinian and Noetherian. Set $M_0 = 0 \subset M$. Since M is Artinian, there exists $M_1 \subset M$ that is minimal among the nonzero submodules of M. Clearly , M_1/M_0 is simple. If $M_1 = M$, then we are done, so assume this is not the case. Again using that M is Artinian, it has a submodule M_2 that is minimal among the nonzero submodules of M properly containing M_1. Then M_2/M_1 is simple: it is nonzero because M_2 properly contains M_1, and if it had a proper nonzero submodule, the inverse image of this submodule in M_2 would properly contain M_1 and be properly contained in M_2, contradicting the minimality of M_2. If $M_2 = M$, then we are done, since then $0 = M_0 \subsetneq M_1 \subsetneq M_2 = M$ is the desired composition series, else define M_3 similarly.

Continuing in this way, we get $M_r = M$ for some r: for otherwise, we would get an infinite strictly ascending chain $0 = M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \dots$, contradicting the assumption that M is Noetherian. Thus, we get a composition series $0 = M_0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_r = M$, so that $l(M) = r$. □

16.3. **Semisimplicity.**

Lemma 16.20. *For a left R-module M, the following are equivalent.*

(SS1) *M is a sum of (a family of) simple left R-modules.*
(SS2) *M is a direct sum of simple left R-modules.*
(SS3) *For every (left R-)submodule $N' \subset M$, there exists a submodule $N \subset M$ such that $N' \oplus N = M$.*

Before we prove the lemma, let us assume it to make the following definition:

Definition 16.21. A left R-module M that satisfies the equivalent conditions of Lemma 16.20 will be called semisimple, or completely reducible.

The proof of Lemma 16.20 will use:

Lemma 16.22. *If an R-module M satisfies (SS3), then every nonzero submodule of M contains a simple submodule.*

Proof. It is clear that any nonzero submodule $M' \subset M$ of M satisfies (SS3): for all $N' \subset M'$ R-submodule, if $N' \oplus N'' = M$, then it is easy to see that $N' \oplus (N'' \cap M') = M'$. Thus, replacing M by any nonzero cyclic (i.e., singly generated) submodule of M', it is enough to start with a nonzero cyclic R-module $M = Rm$ satisfying (SS3), and show that M contains a nonzero simple submodule.

Now that $M = Rm$ is cyclic and nonzero, it is easy to see using a Zorn's lemma argument that it has a maximal proper left R-submodule N: given a chain of proper submodules $\{N_i\}$ of M ordered under inclusion, no N_i contains x, so $N := \bigcup_i N_i$ does not contain x either, and is hence proper.

Now let $N' \subset M$ be such that $M = N \oplus N'$. Then $N' \neq 0$, and it is enough to show that N' is simple. If not, N' contains a proper nonzero R-submodule N'', and $N \oplus N''$ would be a proper submodule of M bigger than N, a contradiction (to put it another, perhaps better, way, $N' \cong M/N$ is simple by the maximality of N). □

Remark 16.23. The end of the argument of the above lemma in fact shows that any finitely generated left R-module M has a maximal proper submodule, and hence a simple quotient. It is (SS3) that allowed us to realize this simple quotient as a simple submodule.

Proof of Lemma 16.20 (some arguments only sketched). (SS1) $\Rightarrow$ (SS3): Let $M = \sum_{i \in I} M_i$, with each M_i a simple left R-module. Let $N \subset M$, and let us find $N' \subset M$ such that $M = N \oplus N'$. Consider the collection of subsets $J \subset I$ such that the sum $N + \sum_{i \in J} M_i \subset M$ is direct: explicitly, those $J \subset I$ such that, whenever $n \in N$ and $(m_i \in M_i)_{i \in J}$ satisfy $n + \sum_i m_i = 0$, we have $n = 0$, and $m_i = 0$ for all $i \in J$. A Zorn's lemma argument, applied to this collection, partially ordered under inclusion, shows that there is a maximal $J \subset M$ such that $N + \sum_{i \in J} M_i = N \oplus (\bigoplus_{i \in J} M_i)$ inside M (this collection is nonempty, because the empty set belongs to it). It is enough to show that the inclusion $N + \sum_{i \in J} M_i \subset M$ is

an equality. If not, there exists $i_0 \in I$ such that $M_{i_0} \nsubseteq (N + \sum_{i\in J} M_i)$. Since M_{i_0} is simple, it follows that $M_{i_0} \cap (N + \sum_{i\in J} M_i) = 0$. But this implies that the sum $N + \sum_{i\in J\cup\{i_0\}} M_i$ is direct. Since clearly $i_0 \notin J$, this contradicts the maximality of J, giving (SS2).

(SS1) $\Leftrightarrow$ (SS2): It is immediate that (SS2) implies (SS1), and the implication (SS1) $\Rightarrow$ (SS2) follows from the exact same argument that was used to prove (SS1) $\Rightarrow$ (SS3), but applied with $N = 0$.

(SS3) $\Rightarrow$ (SS1): Let $N \subset M$ be the sum of all the simple (left R-)submodules of M. It is enough to show that $N = M$. If not, use (SS3) to get $M = N \oplus N'$ for some left R-submodule $N' \subset M$. By Lemma 16.22, there exists a simple submodule $N'' \subset N'$. But by the construction of N, we have $N'' \subset N$, so that $N'' \subset N \cap N' = 0$, a contradiction. □

Corollary 16.24. *The collection of semisimple left R-modules is closed under taking arbitrary direct sums, quotients and submodules. In particular, if R is semisimple as a left R-module, then every R-module is semisimple.*

Proof. Use the criterion (SS1) to prove the assertions about direct sums and quotients. To prove the assertion about submodules, use (SS3): it was observed in the proof of Lemma 16.22 that the condition (SS3) passes to each submodule. For the last assertion, use that every left R-module is a quotient of a free left R-module. □

Theorem 16.25. *Given a ring R, the following are equivalent:*

(i) Every short exact sequence of left R-modules splits.
(ii) Every left R-module is semisimple.
(iii) Every finitely generated left R-module is semisimple.
(iv) Every cyclic (i.e., singly generated) left R-module is semisimple.
(v) R as a left R-module is semisimple.

Proof. (i) is just the assertion that every left R-module M satisfies the condition (SS3) from Lemma 16.20. Therefore, that lemma gives (i) $\Leftrightarrow$ (ii). The implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) are trivial (each is a special case of the previous), while (v) $\Rightarrow$ (ii) follows from Corollary 16.24. □

Definition 16.26. A left semisimple ring is a ring R that satisfies the equivalent conditions of Theorem 16.25. As per our conventions, a right semisimple ring is defined analogously. Later we will hopefully show that, at least in Artinian contexts which will be the ones that we care about, a left semisimple ring is also right semisimple, keeping which in mind we will abbreviate left semisimple to semisimple from now on.

Example 16.27. (i) A field is a semisimple ring.

(ii) We claim that so is a division ring D, i.e., a ring $R = D$ such that each $0 \neq r \in R$ is (both left and right) invertible. In this case, a left or right module over D is also called a left or right D-vector space. D has no proper nonzero left (or right) ideal, and each cyclic left D-vector space is isomorphic to D as a left module over D (use Exercise 16.2). This also implies that this D-vector space is simple, and that every

left vector space over D is a sum of copies of the left D-vector space D, and hence satisfies (SS1). Since (SS1) is equivalent to (SS2) (Lemma 16.20), we have shown that every left D-vector space has a basis (is free). Of course, similarly with right D-vector spaces.

I haven't addressed the question that was asked about the cardinality of the bases, but we will only be interested in finite dimensional D-vector spaces, in which case this cardinality is just the length and hence depends only on the vector space.

(iii) We will hopefully show that if D is a division ring, the matrix algebra $M_n(D)$ is semisimple for each $n \in \mathbb{N}_{\geqslant 1}$.

Lemma 16.28. *(i) A ring R is semisimple if and only if, as a left R-module, R is a direct sum of finitely many of its minimal left ideals.*

(ii) A product of finitely many semisimple rings is semisimple.

Proof. Since minimal left ideals of R are simple modules, and in fact precisely the simple left R-submodules of R, it follows that R is semisimple if and only if, as a left R-module, it is a direct sum of its left ideals. It now suffices to show that, given any decomposition $R = \bigoplus_{i \in J} I_i$ of the left R-module R as a direct sum of nonzero left ideals I_i of R, J is finite. Indeed, write $1 = \sum_{i \in J} x_i$, with $x_i \in I_i$ for each i. This is supported on a finite set $J_0 \subset J$, by the definition of a direct sum. Thus, $R = R \cdot 1 = \sum_{i \in J_0} R \cdot x_i \subset \sum_{i \in J_0} I_j$, so $J = J_0$ is finite (since each I_i was nonzero by assumption).

This gives (i), of which (ii) is an easy corollary: If $R_1 = \bigoplus_i I_{1,i}$ and $R_2 = \bigoplus_j I_{2,j}$, with each $I_{1,i} \subset R_1$ and $I_{2,j} \subset R_2$ a minimal left ideal, then note that as a left module over itself,

$$R_1 \times R_2 = R_1 \times 0 \oplus 0 \times R_2 = \Big(\bigoplus_i I_{1,i} \times 0\Big) \oplus \Big(\bigoplus_j 0 \times I_{2,j}\Big),$$

and note that if $I_1 \subset R_1$ and $I_2 \subset R_2$ are minimal left ideals, then $I_1 \times 0$ and $0 \times I_2$ are minimal left ideals of $R_1 \times R_2$ (in fact these give all the minimal left ideals of $R_1 \times R_2$, but we do not need that). $\square$

16.4. **Jacobson radical.** The Jacobson radical measures one obstruction to R being semisimple; in fact, in the special case where R is Artinian, we will hopefully see that this is the only obstruction.

Its definition can be motivated by the following observation: if R is semisimple, then every $x \in R$ that annihilates every simple (and hence every semisimple) left R-module, annihilates every R-module, and in particular R, and hence vanishes.

On the other hand, we have:

Lemma 16.29. *For $x \in R$, the following are equivalent:*

(i) x annihilates every simple left R-module.

(ii) x is contained in every maximal left-ideal of R.

Proof. By Exercise 16.2, though the annihilator of a simple left R-module may be strictly contained in a maximal left ideal, maximal left ideals are precisely the annihilators of the various nonzero $m \in M$, as M varies over simple left R-modules and m varies over nonzero elements of M. Therefore:

$$\bigcap_{M \text{ simple}} \operatorname{Ann}_R(M) = \bigcap_{\substack{M \text{ simple} \\ 0 \neq m \in M}} \operatorname{Ann}_R(m) = \bigcap_{\mathfrak{m} \text{ maximal left ideal}} \mathfrak{m} = \operatorname{rad}(R).$$

□

Definition 16.30. The Jacobson radical of R, denoted $\operatorname{rad}(R)$, is the collection of all $x \in R$ satisfying the equivalent conditions of Lemma 16.29. Note that $\operatorname{rad}(R)$ is a two-sided ideal of R, since $\operatorname{Ann}_R(M) \subset R$ is a two-sided ideal for each R-module M (easy exercise).

Example 16.31. $\operatorname{rad}(\mathbb{Z}) = 0$, $\operatorname{rad}(\mathbb{Z}_p) = (p)$, $\operatorname{rad}(R/p^n) = (p)$ whenever R is a (commutative) PID, $p \in R$ is prime, and $n \in \mathbb{N}_{\geqslant 1}$. This is generalized by the observation that if R is a commutative ring, and is local, i.e., (in the commutative context) has a unique maximal ideal $\mathfrak{m}$, then $\mathfrak{m}$ is the Jacobson radical of R. More examples can be found in Exercise 16.38 below.

Exercise 16.32. Show that arbitrary ring homomorphisms $R \to S$ do not induce homomorphisms $\operatorname{rad}(R) \to \operatorname{rad}(S)$, but surjective ones do.

Here is another description of $\operatorname{rad} R$:

Lemma 16.33. $\operatorname{rad} R = \{x \in R \mid 1 - yx \text{ is left invertible for all } y \in R\}$.

Remark 16.34. Given the statement of the above lemma, and for later use, let us review left and right invertibility in a ring.

(i) Note that for a general noncommutative ring R, left and right invertibility are different: consider the left and right shift operators in the endomorphism ring of the vector space of all real $\mathbb{N}$-indexed sequences: the left shift operator, which sends $(a_0, a_1, \dots)$ to $(a_1, a_2, \dots)$, is a left inverse to the right shift operator, which sends $(a_0, a_1, \dots)$ to $(0, a_0, a_1, \dots)$. But the left shift operator does not have a left inverse.

(ii) If $x \in R$ has a left inverse u and a right inverse w, then $u = w$, so $u = w$ is a two-sided inverse of x: $w = (ux)w = u(xw) = u$.

(iii) Further, if $x \in R$ has a left inverse u, to show that x also has a right inverse, it is enough to show that u has a left inverse w. Indeed, this follows from:

$$w = w(ux) = (wu)x = x,$$

so that $x = w$ is also a left inverse of u.

Proof of Lemma 16.33. For $x \in R$ and a maximal left ideal $\mathfrak{m} \subset R$, note that

$$x \in \mathfrak{m} \iff Rx + \mathfrak{m} \neq R \iff 1 \notin Rx + \mathfrak{m} \iff (1 - Rx) \cap \mathfrak{m} = \varnothing \iff 1 - Rx \subset R \backslash \mathfrak{m}$$

(for the first and the second steps, use that $Rx + \mathfrak{m}$ is a left ideal containing $\mathfrak{m}$).

Therefore, varying $\mathfrak{m}$ over the maximal left ideals of R, we get

$$\operatorname{rad} R = \left\{ x \in R \mid 1 - Rx \subset R \backslash \left(\bigcup_{\mathfrak{m}} \mathfrak{m} \right) \right\}.$$

Now note that $R\backslash \bigcup_{\mathfrak{m}} \mathfrak{m}$ is simply the set of left invertible elements: $z \in R$ is left invertible if and only if it does not belong to any maximal left ideal. □

Lemma 16.35. $\operatorname{rad} R = \{x \in R \mid 1 - yxz$ *is both left and right invertible for all* $y, z \in R\}$.

Proof. The inclusion "$\supset$" is immediate from Lemma 16.33. Thus, given $x \in \operatorname{rad} R$, it is enough to show that $1 - yxz$ is both left and right invertible for all $y \in R$. Since $\operatorname{rad}(R)$ is a two-sided ideal, we may replace x by yxz: thus, it now suffices to show that for $x \in \operatorname{rad}(R)$, $1 - x$ is both left and right invertible.

By Lemma 16.33, we know that $1 - x$ has a left inverse, say u. To show that $1 - x$ has a right inverse, it suffices, by Remark 16.34(iii), to show that u has a left inverse w. By Lemma 16.33 again, this in turn follows if we show that $u \in 1 - \operatorname{rad}(R)$. For this, note that $u(1 - x) = 1$, so $1 - u = -ux \in \operatorname{rad}(R)$. □

Recall that the ring R^{op} opposite to R has $(R, +)$ for its underlying additive abelian group, but has the multiplication rule switched: the product of x and y in R^{op} is the product of y and x in R (in that order).

Corollary 16.36. $\operatorname{rad}(R) = \operatorname{rad}(R^{op})$, *and* $\operatorname{rad}(R)$ *is the intersection of all the maximal right ideals of* R, *and*

$$\operatorname{rad}(R) = \{x \in R \mid 1 - xy \text{ is right invertible for all } y \in R\}.$$

Proof. Use that the condition in Lemma 16.35 is the same for R and R^{op}, and that maximal left ideals of R identify with maximal right ideals for R^{op}. □

Exercise 16.37. Lemma 16.35 might motivate the question: if $x, y \in R$ are such that $1 + xy$ is invertible, then is $1 + yx$ invertible? Prove that this is so.
Hint: First find a fake proof using binomial expansion, and use the fake proof to guess a recipe for an inverse to $1 + yx$ in terms of an inverse to $1 + xy$.

Exercise 16.38. .

(i) If R is the ring of upper triangular matrices in $M_2(D)$, where D is a division ring, then show as an exercise that $\operatorname{rad}(R)$ is the ideal of strictly upper triangular matrices in R.
Hint: Sending $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ to a or c defines surjective homomorphisms $R \to D$. Thus, by Exercise 16.32, $\operatorname{rad}(R)$ is contained in the subgroup of strictly upper triangular matrices. Now use Lemma 16.35.

(ii) Show that for any ring R, $\mathrm{rad}(M_n(R)) = M_n(\mathrm{rad}(R))$.
Hint: [51] Given $A = [a_{ij}] \in \mathrm{rad}(M_n(R))$ and $y \in R$, left and right multiply it with various matrices to get a matrix with ya_{ij} in a diagonal entry and no other nonzero entry. Thus $1 - ya_{ij}$ is left invertible for all $y \in \mathrm{rad}(R)$ and for all i, j. For the other direction, it is enough to show that if $a \in \mathrm{rad}\, R$ and $1 \leqslant i, j \leqslant n$, then the matrix B that is a in the (i, j)-th entry, and 0 elsewhere, belongs to $\mathrm{rad}(M_n(R))$. Moreover, we may assume that $i = j = 1$ (use permutation matrices). Now by Lemma 16.33, we are reduced to proving that a matrix whose entries differ from the identity only in the first column, and by elements of $\mathrm{rad}(R)$, is left invertible. Use row operations. (Note that you cannot use determinants for this problem since R is noncommutative).

[51] Warning: I haven't spent time checking the following approach I am proposing, use it at your own risk.

17. Lecture 17 – Artinian rings

17.1. **Artinian rings.** In Lecture 16, we saw that if R is semisimple, then $\operatorname{rad}(R) = 0$ (see the discussion before Definition 16.30 in Lecture 16). Moreover, if R is semisimple, we saw that we have a decomposition

$$R = \bigoplus_{i=1}^{n} I_i$$

of R as a left R-module, where each $I_i \subset R$ is a minimal left ideal. Thus, R is of finite length as a left R-module, and hence both Artinian and Noetherian as a left R-module.

Theorem 17.1. *A ring R is semisimple if and only it is left Artinian and* $\operatorname{rad}(R) = 0$.

Proof. As recalled above, we have already seen that if R is semisimple, then $\operatorname{rad}(R) = 0$ and R is left Artinian.

Now assume that R is left Artinian and that $\operatorname{rad}(R) = 0$. Since R is left Artinian, is easy to see that $0 = \operatorname{rad}(R) = \bigcap_{i=1}^{n} \mathfrak{m}_i$ for some maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_n \in R$: if not, we can get an infinite descending chain $\mathfrak{m}_1 \supsetneq \mathfrak{m}_1 \cap \mathfrak{m}_2 \supsetneq \mathfrak{m}_1 \cap \mathfrak{m}_2 \cap \mathfrak{m}_3 \supsetneq \ldots$. Then the obvious map

$$R \to \bigoplus_{i=1}^{n} R/\mathfrak{m}_i$$

of left R-modules (the factors on the right-hand side are not rings) is an inclusion of left R-modules. Since the target of this inclusion is semisimple (being a direct sum of simple left R-modules), so is R. □

Thus, if R is left Artinian, then $R/\operatorname{rad}(R)$ is semisimple with the same collection of simple left R-modules as R.

Remark 17.2. One motivation for the following series of exercises comes from representation theory in bad characteristic: the representations of a finite group G over a finite field k will not in general be completely reducible if $\operatorname{char} k | \#G$. Thus, $k[G]$-modules will not be semisimple in general. But then we can consider the maximal semisimple submodule of a given left $k[G]$-module (i.e., the maximal completely reducible subrepresentation of a representation of G over k), and develop from there into a filtration of the given representation, or go the other way round starting from a maximal completely reducible (or semisimple) quotient. See the following exercise for more details.

Exercise 17.3. For a left R-module M, the radical $\operatorname{rad}(M)$ of M is defined to be the intersection of all the maximal proper submodules of M (if M has no maximal proper submodule, then $\operatorname{rad}(M) = M$).

(i) Suppose M is Artinian. Take ideas from the proof of Theorem 17.1 above to show that $M/\operatorname{rad}(M)$ is the maximal semisimple quotient of M – this means that $M/\operatorname{rad}(M)$ is semisimple, and moreover any quotient map $M \twoheadrightarrow N$, with N

semisimple, factors as $M \to M/\operatorname{rad}(M) \to N$ (in particular, a maximal semisimple quotient of M exists when M is Artinian, which is not a priori obvious).

In general, the maximal semisimple quotient of M is called the cosocle of M – it is unique if it exists, but it need not exist outside the case where M is Artinian: show that the $\mathbb{Z}$-module $\mathbb{Z}$ does not have a cosocle.

(ii) Imitate the proof of Theorem 17.1 above to show that the following are equivalent:
(a) M is semisimple and finitely generated.
(b) M is Artinian and $\operatorname{rad}(M) = 0$.
Thus, $\operatorname{rad}(M)$ measures an obstruction to M being semisimple.

(iii) If the ring R and the left R-module M are Artinian, show that $\operatorname{rad}(M) = \operatorname{rad}(R) \cdot M$.
Hint: By (i) above, it is enough to show that $M/(\operatorname{rad}(R) \cdot M)$ is the maximal semisimple quotient of M. $M/(\operatorname{rad}(R) \cdot M)$ is semisimple because its left R-module structure is inflated from a left $(R/\operatorname{rad}(R))$-module structure, and $R/\operatorname{rad}(R)$ is a semisimple ring by Theorem 17.1 above. On the other hand, if M/N is semisimple, then $\operatorname{rad}(R)$ annihilates it, so that $N \supset \operatorname{rad}(R) \cdot M$.

(iv) The socle of M, $\operatorname{soc}(M)$, is the maximal semisimple submodule of M: it is well-defined since it is equivalently defined as the sum of all the simple submodules of M (see the equivalent characterizations of semisimple modules, Lemma 16.20 from Lecture 16). Show that $\mathbb{Z}$ as a $\mathbb{Z}$-module has socle 0. Show also that if M is Artinian, then M has a nonzero socle.

(v) The socle filtration of M is the largest chain of the form $0 \subsetneq M_1 \subsetneq M_2 \subsetneq \dots$ (possibly ending at some M_n), where M_i/M_{i-1} is the socle of M/M_{i-1} for each $i \geqslant 1$, provided M/M_{i-1} has a nonzero socle (if $\operatorname{soc}(M/M_{i-1}) = 0$, the chain terminates at $M_{i-1} \subset M$).

Show that the socle filtration of $\mathbb{Z}$ has only one term, $\{0\}$. On the other hand, if M has finite length, then show that M has a socle filtration $0 \subsetneq M_1 \subsetneq M_2 \subsetneq \cdots \subsetneq M_n = M$, with $n \leqslant l(M)$.
Note: According to some online dictionary, a socle is "a plain low block or plinth serving as a support for a column, urn, statue, etc. or as the foundation of a wall". The idea seems to be that a general R-module is built somehow from its maximal semisimple submodule, with nontrivial extensions involved.

(vi) As defined earlier, the cosocle of M is the maximal semisimple quotient of M, if such a thing exists.

Show that the cosocle of the Artinian $\mathbb{Z}$-module $(\mathbb{Q}/\mathbb{Z})[p^\infty]$ is 0, and that the radical of this module is itself.

(vii) Define what a cosocle filtration $M = M_0 \supsetneq M_1 \supsetneq \dots$ should mean, and show that if M has finite length, M has a cosocle filtration $M = M_0 \supsetneq M_1 \supsetneq M_2 \supsetneq \cdots \supsetneq M_n = 0$.

17.2. **The Hopkins-Levitzki theorem.** In this subsection, we will prove:

Theorem 17.4 (Hopkins-Levitzki)**.** *If R is left Artinian, and M is a finitely generated left R-module, then M is left Noetherian.*

Before getting to the proof, we will make some preparations, before which, we record the following obvious corollary.

Corollary 17.5. *Any left Artinian ring is left Noetherian.*

Example 17.6. The example of the $\mathbb{Z}$-module $(\mathbb{Q}/\mathbb{Z})[p^\infty]$, which is Artinian but not Noetherian, shows that Theorem 17.4 fails without assuming M to be finitely generated.

Remark 17.7. There is a special case where Theorem 17.4 is trivial. If R is semisimple, then it is immediate (easy exercise) that a left R-module M is Artinian if and only if it is of finite length, and that this is so if and only if M is Noetherian. In a sense, the general case will be reduced to the semisimple case below.

The proof of Theorem 17.4 will use Nakayama's lemma, Proposition 17.8 below:

Proposition 17.8 (Nakayama's lemma)**.** *Let M be a left R-module, and let $I \subset \mathrm{rad}(R)$ be a left ideal, such that $IM = M$. Then, under any of the following additional hypotheses, we have $M = 0$:*

(i) M is finitely generated.
(ii) I is a nilpotent ideal, in the sense that some finite product $I \cdot I \cdot \cdots \cdot I = 0$.
(iii) $I \supseteq I^2 \supseteq \ldots$ stabilizes, and M is Artinian.

Before we prove the above lemma, let us note the following corollary:

Corollary 17.9. *Assume that we are given a left R-module M, and a left ideal $I \subset \mathrm{rad}(R)$ such that the condition (i), (ii) or (iii) of Proposition 17.8 is satisfied. Let $N \subset M$ be a submodule such that $M = N + IM$. Then $N = M$.*

Proof. Note that Proposition 17.8 can be applied to M/N. □

Proof of Proposition 17.8. The proof under the assumption (ii) is trivial: $M = IM = I^2M = \ldots$.

Let us prove it under the assumption (i). Suppose M is nonzero. Then M has a maximal proper submodule M_0 (this follows from an easy Zorn's lemma argument, using that M is finitely generated). Therefore M/M_0 is simple, and is hence annihilated by $\mathrm{rad}(R)$. Thus, $\mathrm{rad}(R)M \subset M_0$, so that $M = IM \subset M_0$, a contradiction.

Here is a proof of the same assertion when M is cyclic: in this case, for a generator $m \in M$, $\mathrm{rad}(R)m = \mathrm{rad}(R)Rm = \mathrm{rad}(R)M = M$ (use that $\mathrm{rad}(R)$ is a two-sided ideal), so we can write $m = im$ with $i \in \mathrm{rad}(R)$ and $m \in M$; then $(1-i)m = 0$, and since $1-i$ is invertible, we get $m = 0$. In the commutative case, this can be made into a proof using a finite set of generators and determinants, avoiding Zorn's lemma. But determinants don't work well over noncommutative rings.

Now let us prove that $M = 0$ if (iii) is satisfied. Suppose not. Replacing I by some power I^n such that $I^n = I^{n+1} = \ldots$, we may assume without loss of generality that

$I = I^2 = \dots$. Since M is Artinian and nonzero, there exists $M_0 \subset M$ minimal such that $IM_0 \neq 0$. Choosing $m_0 \in M_0$ such that $Im_0 \neq 0$, we have $I \cdot Im_0 = Im_0 \neq 0$, and Im_0 is a submodule of M, so the minimality of M_0 implies that $Im_0 = M_0$. This implies that $im_0 = m_0$ for some $i \in I$. Since $1 - i$ is invertible, it follows that $m_0 = 0$, a contradiction. □

The following lemma is a very helpful consequence, and is a priori quite nonobvious.

Lemma 17.10. *If R is left Artinian, then* $\mathrm{rad}(R)$ *is nilpotent.*

Proof. Since R is left Artinian, we have $\mathrm{rad}(R)^n = \mathrm{rad}(R)^{n+1} = \dots$ for some n. Let $M = \mathrm{rad}(R)^n$, and let $I = \mathrm{rad}(R)$. Then $I \subset \mathrm{rad}(R), IM = M$, and the condition (iii) of Proposition 17.8 is satisfied. Therefore $M = 0$, i.e., $\mathrm{rad}(R)^n = 0$. □

Now we can prove the theorem of Hopkins and Levitzki:

Proof of Theorem 17.4. Let $J = \mathrm{rad}(R)$, and use Lemma 17.10 to choose n large enough such that $J^n = 0$. Thus, it is enough to show that each $(J^iM)/(J^{i+1}M)$ is Noetherian as a left R-module, or equivalently, as a left R/J-module ($J \subset R$ is a two-sided ideal, so R/J is indeed a ring, and $J \subset R$ clearly annihilates $(J^iM)/(J^{i+1}M)$). In what follows, we will use that R/J is a semisimple ring by Theorem 17.1: to see that this theorem applies, note that:

- R/J is left-Artinian as R is, and that
- $\mathrm{rad}(R/J)$ is trivial (use that the maximal left ideals of R all contain J and are hence in bijection with those of R/J).

Now M, being finitely generated as a left module over the Artinian ring R, is easily seen to be Artinian: use Lemma 16.17 from Lecture 16. Hence, by the same lemma, $J^iM/(J^{i+1}M)$ is an Artinian left module over R, and hence over R/J. Therefore, using that R/J is semisimple, by the already observed semisimple case of the theorem (see Remark 17.7), we have that $J^iM/(J^{i+1}M)$ is a Noetherian left module over R/J, and hence over R, as desired. □

Note that the proof of the above theorem was far from immediate: several distinct ideas went into it, such as Lemma 17.10 which used Proposition 17.8.

17.3. **Locally nilpotent ideals and the Jacobson radical.** When R is commutative, it is easy to see that the set $\mathrm{Nil}(R)$ of nilpotent elements of R is an ideal. It is called the nilradical of R.

Lemma 17.11. *When R is commutative,* $\mathrm{Nil}(R)$ *is the intersection of all prime ideals of R, and hence is contained in* $\mathrm{rad}(R)$.

Proof. If $x^n = 0$ for some $n \in \mathbb{N}$, then for each prime ideal $\mathfrak{p} \subset R$, we have $x^n \in \mathfrak{p}$, so $x \in \mathfrak{p}$. Thus, Nil(R) is contained in the intersection of the prime ideals of R.

To show that the intersection of the prime ideals of R is contained in the nilradical, given $a \in R$ nonnilpotent, it is enough to show that there exists a prime ideal $\mathfrak{p}$ of R that does not contain any power of a. Use Zorn's lemma to show that there exists an ideal $\mathfrak{p} \subset R$ that does not contain any power of a (their collection is nonempty since 0 belongs to this collection – since a is nonnilpotent), and is maximal with respect to this property. If $\mathfrak{p}$ is prime we are done, so suppose not, so $\exists\, x, y \in R \backslash \mathfrak{p}$ such that $xy \in \mathfrak{p}$. Then $\mathfrak{p} + (x)$ and $\mathfrak{p} + (y)$ each contains some power of a, so $\mathfrak{p} \supset (\mathfrak{p} + (x))(\mathfrak{p} + (y))$ itself contains some power of a, a contradiction. □

Remark 17.12. Those of you who have seen localization will recognize what is happening here: if a is nonnilpotent, then the localization R_a of R at the multiplicatively closed subset $\{1, a, a^2, \dots\}$ is nonzero, and all that we have done is to take the inverse image, under $R \to R_a$, of a maximal ideal of R_a: a maximal ideal of R_a does not pullback to a maximal ideal of R, but a prime (and in particular maximal) ideal of R_a does pull back to a prime ideal of R; it is elementary that that is how ring homomorphisms work.

Definition 17.13. An ideal $I \subset R$ is called locally nilpotent if each element of I is nilpotent, i.e., for all $a \in I$, $\exists\, n \in \mathbb{N}$ such that $a^n = 0$.

Lemma 17.14. *If $I \subset R$ is a locally nilpotent left ideal, then $I \subset \mathrm{rad}(R)$.*

Proof. If I is locally nilpotent, then for all $x \in I$ and $y \in R$, yx belongs to I and is hence nilpotent. Therefore $1 - yx$ is left-invertible (in fact left and right invertible). Since this is true for all $y \in R$, we get that $x \in \mathrm{rad}(R)$ by Lemma 16.33 from Lecture 16. □

17.4. **Artin local rings, modulo lifting of idempotents.**

Definition 17.15. A ring is called local if the noninvertible elements in it form a two-sided ideal.

Exercise 17.16. (This exercise will often be implicitly assumed in what follows, including in the subsequent lectures). Show that a commutative ring is local if and only if it has a unique maximal ideal.

Thus, local rings are much simpler than usual rings.

We would like to prove:

Theorem 17.17. *A commutative Artin ring is a finite direct product of Artin local rings.*

Example 17.18. (i) If n is a natural number, then $\mathbb{Z}/n\mathbb{Z}$ is an Artinian ring (being finite), and is local if and only if n is a prime power (very easy exercise). Thus, if $n = \prod_i p_i^{a_i}$ is the prime factorization of n (so the p_i are distinct), then the product

decomposition given by Theorem 17.17 can be realized using Sunzi's theorem, i.e., the Chinese remainder theorem:

$$\mathbb{Z}/n\mathbb{Z} \cong \prod_i \mathbb{Z}/p_i^{a_i}\mathbb{Z}.$$

(ii) For any commutative Noetherian ring R, given a maximal ideal $\mathfrak{m} \subset R$ and an ideal $\mathfrak{q} \subset R$ with $\mathfrak{m}^{n+1} \subset \mathfrak{q} \subset \mathfrak{m}^n$, $R/\mathfrak{q}$ is Artin local: prove this as an exercise.
Hint: Any commutative ring with a nilpotent maximal ideal is local, by Lemma 17.11. Secondly, use that if $\bar{R}$ is Noetherian and $\bar{\mathfrak{m}}^{n+1} = 0$, then R as an R-module has a filtration $\bar{R} \supset \bar{\mathfrak{m}} \supset \bar{\mathfrak{m}}^2 \supset \cdots \supset \bar{\mathfrak{m}}^{n+1} = 0$, with each successive quotient being of the form $\bar{\mathfrak{m}}^i/\bar{\mathfrak{m}}^{i+1}$, which is a finitely generated $\bar{R}/\bar{\mathfrak{m}}$-module by Noetherianness, and hence of finite length ($\bar{R}/\bar{\mathfrak{m}}$ is a field).

One technical input into the proof of Theorem 17.17 is:

Lemma 17.19. *Let R be a (not necessarily commutative) ring, and $I \subset R$ a nilpotent* [52] *two-sided ideal, so that R/I is a ring.*

(i) The reduction map $R \to R/I$ induces a surjection

$$\{\textit{Idempotents in } R\} \to \{\textit{Idempotents in } R/I\}.$$

(ii) If R is commutative, then the surjection of (i) is a bijection.

Before proving either of Lemma 17.19 or Theorem 17.17, let us discuss the relevance of the former to the latter.

Definition 17.20. (1) Two central idempotents $e, e' \in R$ are said to be orthogonal to each other if $ee' = 0$. Note that if $e, e' \in R$ are orthogonal, then $e + e'$ is an idempotent as well.
(2) A collection of mutually orthogonal central idempotents $e_1, \ldots, e_n \in R$ is said to be complete if $e_1 + \cdots + e_n = 1$.

Exercise 17.21. Show that for a not necessarily commutative ring R, and a positive integer n, we have bijections between the following collections of objects:

(i) Decompositions $R = I_1 \oplus \cdots \oplus I_n$ of R as a direct sum of some of its nonzero two-sided ideals;
(ii) Decompositions $R = R_1 \times \cdots \times R_n$ of R as a product of nonzero rings, up to a suitable notion of equivalence; and
(iii) Decompositions $1 = e_1 + \cdots + e_n$ of $1 \in R$ as a sum of a complete set of mutually orthogonal central nonzero idempotents $e_1, \ldots, e_n$,

given as follows:

[52] this lemma is actually true with 'nilpotent' replaced by 'locally nilpotent'; see Professor Nair's notes, or Remark 17.26 below for a discussion.

- Given $R = I_1 \oplus \cdots \oplus I_n$ as in (i), each $I_i \subset R$, with its induced multiplication, is a ring with multiplicative identity the image of $1 \in R$ under the projection $R \to I_i$; writing $R_i = I_i$ for this ring, this gives homomorphisms $R \to R_i$ for $1 \leqslant i \leqslant n$, and an isomorphism $R \to R_1 \times \cdots \times R_n$ of rings as in (ii).
- Given $R = R_1 \times \cdots \times R_n$ as in (ii), for $1 \leqslant i \leqslant n$ let $e_i \in R$ be the unique idempotent whose image in R_j is 0 if $i \neq j$, and 1 if $i = j$; then $(e_1, \ldots, e_n)$ is as in (iii).
- Given $(e_1, \ldots, e_n)$ as in (iii), set $I_i = e_i R = Re_i \subset R$, which can also be viewed as a ring R_i; then the projection $R \to I_i$ is a ring homomorphism $R \to R_i$, and we get a decomposition $R = I_1 \oplus \cdots \oplus I_n$ as in (i) as well as a decomposition $R = R_1 \times \cdots \times R_n$ as in (ii). Note that we can thus also write $R = Re_1 \times \cdots \times Re_n$.

The above exercise is implicitly used often in representation theory: the set of simple left modules over $R = R_1 \times \cdots \times R_n$ is a disjoint union of the sets of simple left modules over the R_i (a more detailed formulation will be given in an exercise in Lecture 18). Thus, finding central idempotents in R lets us break down the representation theory of R into that of 'smaller' and hopefully easier rings.

Now let us use Lemma 17.19 to prove Theorem 17.17.

Proof of Theorem 17.17. Since R is Artinian, there are finitely many maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_n \subset R$ such that

$$\operatorname{rad}(R) = \bigcap_{i=1}^{n} \mathfrak{m}_i.$$

Since $R/\operatorname{rad}(R)$ is Artinian and semisimple (and commutative), we get by Sunzi's theorem a product decomposition

$$R/\operatorname{rad}(R) \cong \prod_{i=1}^{n} (R/\mathfrak{m}_i).$$

This decomposition corresponds to a complete set $\bar{e}_1, \ldots, \bar{e}_n \in R/\operatorname{rad}(R)$ of pairwise orthogonal nonzero idempotents, which by Lemma 17.19(ii) can be lifted to unique idempotents $e_1, \ldots, e_n \in R$: note that Lemma 17.19 applies since $\operatorname{rad}(R)$ is nilpotent by Lemma 17.10.

We claim that $e_1, \ldots, e_n$ is a complete set of pairwise orthogonal idempotents in R. Since $e_i e_j$ is an idempotent lifting $\bar{e}_i \bar{e}_j = 0$, by Lemma 17.19(ii) we get $e_i e_j = 0$ for $i \neq j$. This gives the pairwise orthogonality and also implies that $e_1 + \cdots + e_n$ is an idempotent; since this idempotent lifts $1 \in R/\operatorname{rad}(R)$, we get that $e_1 + \cdots + e_n = 1$.

Again by Exercise 17.21, this gives us a decomposition $R \cong \prod_{i=1}^{n} Re_i$. Clearly each Re_i is Artinian (being a quotient of R), and it remains to show that each Re_i is local.

$R \to R/\operatorname{rad}(R)$ induces a ring homomorphism, $Re_i \to (R/\operatorname{rad}(R))\bar{e}_i \cong R/\mathfrak{m}_i$, which, viewed as an R-module homomorphism, is also a restriction of $R \to R/\operatorname{rad}(R)$. Thus, the kernel of $Re_i \to (R/\operatorname{rad}(R))\bar{e}_i \cong R/\mathfrak{m}_i$ is contained in $\operatorname{rad} R$, which is nilpotent (Lemma 17.10). But the image of this ring homomorphism is a field, $R/\mathfrak{m}_i$. Thus, this kernel is both

nilpotent and a maximal ideal. But by Lemma 17.14, this kernel is contained in $\mathrm{rad}(Re_i)$, and hence by maximality also the unique maximal ideal of Re_i. Therefore Re_i is local. □

Exercise 17.22. (i) In the setting of the above proof, show that each maximal ideal of R is of the form $\mathfrak{m}_i$ for some $1 \leqslant i \leqslant n$. Thus, a commutative Artinian ring has only finitely many maximal ideals (though a semisimple module that is Artinian and hence of finite length can have infinitely many maximal proper submodules).

(ii) The above proof simplifies in the case where R is semisimple, since in this case $R = R/\mathrm{rad}(R)$ – note that, in this case, there is no need for Lemma 17.19. Using this, show that a commutative ring R is semisimple if and only if it is a direct product of fields (recall that a semisimple ring is automatically Artinian). Can you show directly that a product of fields is semisimple?

17.5. **Hensel's lemma and lifting idempotents.** To complete the proof of Theorem 17.17, we need to prove the lemma on lifting idempotents, Lemma 17.19. For this, we will use a very important result called Hensel's lemma, which is quite widely used in algebraic number theory and algebraic geometry.

Theorem 17.23 (A variant of Hensel's lemma)**.** *Let R be a commutative ring with a nilpotent ideal I. Let $a \mapsto \bar{a}$ denote reduction modulo I, at the level of polynomials as well. Let f be a polynomial in $R[x]$ such that $\bar{f} \in R[x]/(IR[x]) = \bar{R}[x]$ has a root $\bar{\alpha} \in \bar{R}$, and suppose $\bar{f}'(\bar{\alpha}) \in \bar{R}^\times$. Then $\bar{\alpha}$ can be lifted to a unique root $\alpha \in R$ of f.*

The proof of Theorem 17.23 will in turn use the following easy observation:

Lemma 17.24. *If R is a commutative ring and $I \subset R$ is a nilpotent ideal, then $R^\times$ is the full preimage of $(R/I)^\times$ under $R \to R/I$.*

Proof. It is immediate that $R^\times$ maps to $(R/I)^\times$. On the other hand, if $a \in R$ is invertible modulo I, so that $ab \in 1 + I$ for some $b \in R$, then since $1 + I$ consists of units, it follows that $a \in R^\times$, as desired. □

Proof of Theorem 17.23. The proof is by Hensel/Newton/... depending on how you interpret it. First we prove the existence. Let $\alpha_1 \in R$ be any lift of $\bar{\alpha}$. Since $\bar{f}(\bar{\alpha}) = 0$, we have $f(\alpha_1) \in I$. Inductively, we will construct $\alpha_2, \alpha_3, \ldots$, all reducing to $\bar{\alpha} = \bar{\alpha}_1$ modulo I, such that $f(\alpha_n) \in I^n$ for each n: this will suffice for the existence assertion, since I is nilpotent. We already have constructed α_1; assume that we have constructed $\alpha_1, \ldots, \alpha_n$.

Note that $\overline{f'(\alpha_n)} = \bar{f}'(\bar{\alpha}) \in (R/I)^\times$. By Lemma 17.24, we have $f'(\alpha_n) \in R^\times$.

Thus, given $\beta \in R$ such that $f(\beta) \in I^n$ and $f'(\beta) \in R^\times$, it is enough for the existence assertion to show that there exists $\gamma \in R$ such that $f(\gamma) \in I^{n+1}$. In fact, take $\gamma = \beta - \frac{f(\beta)}{f'(\beta)}$ (see the "Newton" angle here: Newton's method of locating roots of a polynomial etc.). Since f is a polynomial, we have

$$f(x + h) = f(x) + f'(x)h + (\text{higher degree terms in } h),$$

by the binomial theorem (e.g., see it degree by degree). Apply this with $x = \beta$ and $h = \gamma - \beta \in I^n$. Each higher degree term in h belongs to $I^{2n} \subset I^{n+1}$. Therefore, we get

$$f(\gamma) = f(\beta) - f(\beta) + I^{n+1} = I^{n+1},$$

as desired.

Now we come to the uniqueness. If distinct α, β both lift $\bar{\alpha}$ and satisfy $f(\alpha) = f(\beta) = 0$, then let n be the unique positive integer such that $\beta - \alpha \in I^n \backslash I^{n+1}$. Applying the same Taylor expansion for $f(x + h)$ as above, this time with $x = \alpha$ and $h = \beta - \alpha$, we get

$$0 = f(\beta) - f(\alpha) \in f'(\alpha)(\beta - \alpha) + I^{n+1} \not\equiv 0 \bmod I^{n+1},$$

a contradiction. □

Exercise 17.25. (i) Extend Theorem 17.23 to the case where R is complete with respect to I, i.e., where the obvious map

$$R \to \varprojlim_n R/I^n,$$

i.e., the map obtained from the projections $R \to R/I^n$, is an isomorphism (But of course without assuming that I is nilpotent).
Hint: This is pretty much all done in the proof of Theorem 17.23: defining α_n for each n exactly as in that proof, we have $(\alpha_n)_n \in \varprojlim_n R/I^n$ is a root of f in the R-algebra $\varprojlim_n R/I^n$.

(ii) Let p be a prime number. Show that $\mathbb{Z}_p$ is a local ring, and that it contains $p - 1$ $(p-1)^{\text{th}}$ roots of unity.
Hint: $\mathbb{F}_p^\times$ contains $p - 1$ $(p-1)^{\text{th}}$ roots of unity.

Now we use Theorem 17.23 to prove Lemma 17.19, completing the proof of Theorem 17.17.

Proof of Lemma 17.19. To prove (i), we may replace R with the subring $R_0 \subset R$ generated by the image of $\mathbb{Z} \to R$ together with any chosen lift of a given idempotent $\bar{e}$ in R/I, and I with $I \cap R_0$, to assume that R is commutative. Thus, we may assume that R is commutative, and prove just (ii) of the lemma.

We wish to apply Theorem 17.23 with $f(x) = x^2 - x$. To show that $\bar{e}$ has a unique idempotent lift in R, it is enough to show that $\bar{f}'(\bar{e}) \in (R/I)^\times$. In other words, it is enough to show that $2\bar{e} - 1$ is a unit in $\bar{R} := R/I$. But this follows from the fact that $(2\bar{e} - 1)^2 = 4\bar{e} - 4\bar{e} + 1 = 0$.

If squaring $2\bar{e} - 1$ seems unmotivated, note that $\bar{e}$ gives a decomposition $\bar{R} = \bar{R}_1 \times \bar{R}_2$, under which $2\bar{e} - 1$ corresponds to $(1, -1) \in \bar{R}_1^\times \times \bar{R}_2^\times \cong \bar{R}^\times$, which has square 1. □

Remark 17.26. To apply Theorem 17.23, we need I to be nilpotent, and not just locally nilpotent. However, at least if we allow ourselves to use the Hilbert basis theorem, one can show that after replacing R with R_0 and I with $I \cap R_0$ as above, I becomes finitely generated; a finitely generated ideal that is locally nilpotent is nilpotent. For the proof of

uniqueness too, with a bit of work one can pass to the case of I being nilpotent. For a direct argument which doesn't need such reductions, see Professor Nair's notes; I avoided that argument since I wanted to take this opportunity to introduce Hensel's lemma.

In any case, for our purposes here, which was for the application of Lemma 17.19 to Theorem 17.17, it sufficed to work with I nilpotent, since $\mathrm{rad}(R)$ is nilpotent.

18. Lecture 18 – Indecomposable modules, the Krull-Schmidt-Remak theorem, Artin-Wedderburn theorem

18.1. **Indecomposable modules and the Krull-Schmidt-Remak theorem.**

Definition 18.1. A left R-module M is called indecomposable if it is nonzero and cannot be written as the direct sum of two of its nonzero submodules.

Remark 18.2. If R is semisimple, then indecomposable modules over R are the same as the simple modules over R, but not in general.

Exercise 18.3. Let G be a cyclic group of order n, and k an algebraically closed field of characteristic p. Recall $Rep_k(G)$ is identified with $k[G]$-Mod. Note that choosing a generator of G gives an isomorphism $k[G] \cong k[T]/(T^n - 1)$.

(i) In good characteristic, i.e., when $(n, p) = 1$, show that $k[G]$ is semisimple. Give a bijection

$$\left\{ \begin{array}{c} \text{the set of isomorphism classes} \\ \text{of simple (eqiv., indecomposable)} \\ k[G]\text{-modules of } G \end{array} \right\} \to \mu_n(k),$$

where $\mu_n(k)$ is the set of n-th roots of unity in k, which are n in number.

(ii) Consider the opposite case, i.e., when n is a p-power. Show that there is only one simple module of $k[G]$ up to isomorphism, given by the trivial representation of G over k, but that there are still exactly n indecomposable $k[G]$-modules up to isomorphism, namely the $k[T]/(T^m)$ with $1 \leqslant m \leqslant n$.
Hint: One way to do this is to use the Jordan canonical form, another is to appeal to HW 1, where you classified indecomposable modules over a PID (which $k[T]$ is).

(iii) Combine the two extreme cases to show that, no matter what n is, there are exactly n indecomposable representations of G up to isomorphism.

Exercise 18.4. Read up about uniserial modules – those for which the successive quotients of the socle filtration (or, equivalently as it turns out, the cosocle filtration) are actually simple (and not just semisimple). Using your work for Exercise 18.3 above, show that that indecomposable representations of finite cyclic groups (in arbitrary characteristic) are uniserial, something that is not true for more general modules.

The main result we would like to prove regarding indecomposable modules is the Krull-Schmidt-Remak decomposition:

Theorem 18.5 (Krull-Schmidt-Remak decomposition)**.** *(i) If M is a left R-module of finite length, then it has a direct sum decomposition $M = \bigoplus_{i=1}^{r} U_i$, with each submodule $U_i \subset M$ an indecomposable left R-module.*

(ii) If $M = \bigoplus_{i=1}^{r} U_i$ and $M = \bigoplus_{i=1}^{s} V_i$ are decompositions of M into indecomposable modules, then $r = s$, and after permuting the $\{V_i\}$ if necessary we have $U_i \cong V_i$ for each i.

Example 18.6. Over a (commutative) PID R, you proved in HW 1 that the only indecomposable modules are R and those of the form $R/(p^i)$, where $p \in R$ is prime and $i \in \mathbb{N}_{\geqslant 1}$. This used the existence assertion of the structure theorem for modules over a PID, but not the uniqueness assertion. Combining this with the uniqueness assertion in the Krull-Schmidt-Remak decomposition (Theorem 18.5(ii)), one can deduce the uniqueness assertion in the structure theorem for modules over a PID: see the remarks made in the notes for Lecture 2. Moreover, we can eliminate the dependence on the proof of the existence assertion given in Lecture 2: one way to prove that every indecomposable torsion module over the PID R is of the form $R/(p^i)$ to use Baer's criterion, as we saw in Lecture 14.

Remark 18.7. The Krull-Schmidt-Remak theorem seems to be of a very different nature from observations regarding the socle and the cosocle filtrations. According to an analogy I saw in the book 'Local Representation Theory' by Alperin (assuming I understand it correctly), if a module is thought of as a cake, the socle/cosocle filtrations can be thought of as slicing the cake horizontally, while the Krull-Schmidt-Remak theorem is then like slicing it vertically.

The main input we will use for the proof of Theorem 18.5 is:

Lemma 18.8. *For a left R-module M of finite length, the following are equivalent:*

(i) M is indecomposable.
(ii) The ring $\mathrm{End}_R(M)$ is local (recall that this means that the set of noninvertible elements in it form a two-sided ideal). [53]

Compare the above lemma with the following lemma, Schur's lemma, which is immediate (and yet very important):

Lemma 18.9. *For a left R-module M, the first of the following two conditions implies the second:*

(i) M is simple.
(ii) The ring $\mathrm{End}_R(M)$ is a division ring (which automatically makes it local).

The following lemma is actually much simpler than Lemma 18.8, and yet seems to have vague parallels with; our awkward statement intends to highlight this parallel.

Lemma 18.10. *Let M, N be left modules over R, with M nonzero and N indecomposable. Then given any $a \in \mathrm{Hom}_R(M, N)$, there are only the following two possibilities:*

- *a an isomorphism, or*
- *For all $s \in \mathrm{Hom}_R(N, M)$, $s \circ a \in \mathrm{End}_R(M)$ is not an isomorphism.*

(This is just a convoluted way of saying that if $s \circ a$ is an isomorphism, then so is a).

[53] Recall that invertible means 'both left and right invertible', and that if an element of R has both a left inverse and a right inverse, these inverses coincide.

Proof. Suppose that $a \in \mathrm{Hom}_R(M,N)$ and $s \in \mathrm{Hom}_R(N,M)$ are such that $s \circ a \in \mathrm{End}_R(M)$ is an isomorphism. It is enough to show that a is an isomorphism. Changing s if necessary, we may and do assume that $s \circ a = \mathrm{id}_M$.

It is easy to see that $N \cong a(M) \oplus \ker s$: the obvious map $a(M) \oplus \ker s \to N$, given by addition in N, has a two-sided inverse given by $n \mapsto (a(s(n)), n - a(s(n)))$.

Since N is indecomposable, and since $a(M)$ is nonzero (as it has the left inverse s, and since $M \neq 0$), this forces $\ker s = 0$, so that $a : M \to N$ is an isomorphism as desired.

□

Let us now use Lemma 18.8 to prove Theorem 18.5:

Proof of Theorem 18.5, assuming Lemma 18.8. The existence is immediate from induction and the finite length assumption. So let us prove the uniqueness, which is the nontrivial part.

We will show that, after permuting the V_i if necessary, we can ensure that

$$a_1 : U_1 \hookrightarrow M = \bigoplus_{i=1}^{s} V_i \to V_1 \tag{74}$$

is an isomorphism. If this is granted, then $U_1 \hookrightarrow M \to M/(\bigoplus_{i=2}^{s} V_i)$ is an isomorphism, so $M = U_1 \oplus (\bigoplus_{i=2}^{s} V_i)$, and it then suffices to apply the induction hypothesis to

$$M/U_1 \cong \bigoplus_{i=2}^{r} U_i \cong \bigoplus_{i=2}^{s} V_i.$$

Thus, let us prove that after some permutation (if necessary) of the V_i's, a_1 becomes an isomorphism.

Consider, for each $1 \leqslant i \leqslant s$,

$$\phi_i : U_1 \hookrightarrow M = \bigoplus_{i=1}^{s} V_i \to V_i \hookrightarrow M = \bigoplus_{i=1}^{r} U_i \to U_1,$$

where each map is an obvious inclusion or a projection.

Clearly, $\sum_{i=1}^{s} \phi_i = \mathrm{id}_{U_1} \in \mathrm{End}_R(U_1)$. Since $\mathrm{End}_R(U_1)$ is local by Lemma 18.8, it follows that not all the ϕ_i can be noninvertible (else so would be their sum, id_{U_i}), so some $\phi_i : U_1 \to U_1$ is an isomorphism. After permuting the V_i, we may and do assume that ϕ_1, $U_1 \to V_1 \to U_1$, is an isomorphism.

Therefore, by Lemma 18.10 (using that V_1 is indecomposable), $U_1 \to V_1$ is an isomorphism.

□

The proof of Lemma 18.8 will in turn be based on the following two results:

Lemma 18.11. *If all the noninvertible elements of R are nilpotent, then R is local.*

Lemma 18.12 (Fitting's lemma)**.** *Let M be a left R-module.*

(i) *If M is Artinian, there exists $n_0 \in \mathbb{N}$ such that $\ker(u^n)+\mathrm{im}(u^n) = M$ for all $n > n_0$.*
(ii) *If M is Noetherian, there exists $n_0 \in \mathbb{N}$ such that $\ker(u^n) \cap \mathrm{im}(u^n) = 0$ for all $n > n_0$.*
(iii) *If M is of finite length (and hence Artinian and Noetherian), there exists $n_0 \in \mathbb{N}$ such that $M = \ker(u^n) \oplus \mathrm{im}(u^n)$ for all $n > n_0$.*

Let us prove Lemma 18.8 assuming Lemmas 18.11 and 18.12.

Proof of Lemma 18.8. If M is not indecomposable, then $\mathrm{End}_R(M)$ has idempotents e_1, e_2 with $e_1 + e_2 = 1$, so the noninvertible elements $e_1, e_2 \in R$ cannot both belong to a proper ideal of R. Thus, it suffices to assume that M is indecomposable, and show $\mathrm{End}_R(M)$ to be local.

Since M is of finite length, Lemma 18.12 implies that given any $u \in \mathrm{End}_R(M)$, for large n we have that $M = \ker(u^n) \oplus \mathrm{im}(u^n)$. By indecomposability, for large n either $\ker(u^n) = 0$ and $\mathrm{im}(u^n) = M$, or $\ker(u^n) = M$ and $\mathrm{im}(u^n) = 0$. In the former case u^n is invertible and hence so is u, while in the latter case u is nilpotent. Thus, by Lemma 18.11, $\mathrm{End}_R(M)$ is a local ring, as desired. □

To complete the proof of Lemma 18.8 and hence of Theorem 18.5, it remains to prove Lemmas 18.11 and Lemma 18.12.

Proof of Lemma 18.11. We claim that the set of noninvertible elements in R is closed under left and right multiplication. If $x \in R$ is noninvertible and hence nilpotent, and $y \in R$, then for some n we have $x^{n-1} \neq 0$ but $x^{n-1} \cdot xy = 0$ and $yx \cdot x^{n-1} = 0$, so that neither xy nor yx is invertible.

It now suffices to assume that $x, y \in R$ are nilpotent and show that $x + y$ is nilpotent as well. Indeed, otherwise there exists $r \in R$ with $r(x + y) = 1$, so that $ry = 1 - rx$ is a unit (since rx is nilpotent by the above paragraph), which contradicts the above paragraph. □

Proof of Lemma 18.12. (iii) follows from (i) and (ii).

For (i), if M is Artinian, then the chain $\mathrm{im}(u) \supseteq \mathrm{im}(u^2) \supseteq \ldots$ stabilizes, say at $\mathrm{im}(u^{n_0})$. Thus, for $m \in M$ and $n > n_0$, we have $u^n(m) = u^{2n}(m')$ for some m', so that $m - u^n(m') \in \ker(u^n)$. Thus, $M = \ker(u^n) + \mathrm{im}(u^n)$ for all $n > n_0$, as desired.

For (ii), if M is Noetherian, then the chain $\ker(u) \subseteq \ker(u^2) \subseteq \ldots$ stabilizes, say at $\ker(u^{n_0})$. Thus, if $n > n_0$ and $m \in \ker(u^n) \cap \mathrm{im}(u^n)$, say $m = u^n(m')$, then $u^n(m) = u^{2n}(m') = 0$, so $m' \in \ker(u^{2n}) = \ker(u^n)$, so $m = u^n(m') = 0$, as desired. □

Remark 18.13. (Optional) In a previous lecture, I remarked that a good chunk of what we are studying currently adapts immediately to a general abelian category. But the Krull-Schmidt-Remak theorem does not, since the arguments in the proof of Lemma 18.12 are specific to the category *R-Mod* (we worked with specific elements of modules), and Lemma 18.12 was in turn crucially used in the proof of the main technical input into the proof of

Theorem 18.5, namely, the assertion that a finite length object of R-Mod is indecomposable if and only if its endomorphism ring is local (i.e., Lemma 18.11).

However, there are abelian categories with additional properties where this can be made to work, as I now summarize mostly from Professor Nair's notes, to which I refer for a more detailed discussion.

- One can consider those abelian categories where objects A in an abelian category $\mathcal{A}$ satisfy a certain 'bi-chain condition', which ensures that an object is indecomposable if and only if its endomorphism ring is local (i.e., the analogue of the key Lemma 18.11 holds).
- One way to ensure that each object satisfies this bi-chain condition is to consider what are known as Hom-finite abelian categories: here one considers not just abelian categories, but k-linear abelian categories $\mathcal{A}$, namely one where the Hom's are not just abelian groups, but modules over a commutative ring k; the condition is then that $\mathrm{Hom}_{\mathcal{A}}(X, Y)$ is of finite length as a k-module for each X, Y (in a manner respecting composition).
- One can show that the bi-chain condition is satisfied by any object in the category of coherent sheaves on a complete variety over an algebraically closed field, and by any object in the category of coherent analytic sheaves on a compact complex manifold.
- In general, there is also such a thing as a Krull-Schmidt category, a category where an appropriate analogue of the Krull-Schmidt-Remak theorem holds (e.g., see wikipedia). The previous point says that the category of coherent sheaves on a complete variety over an algebraically closed field is a Krull-Schmidt category, as is the category of coherent analytic sheaves on a compact complex manifold.

18.2. **Another proof of the structure theorem for commutative Artin rings.** We can now get another proof of the assertion that any commutative Artin ring is a product of Artin local rings:

Another proof of Theorem 17.17 from Lecture 17. R as an R-module is Artinian and Noetherian, the latter because of the Hopkins-Levitzki theorem, and is hence of finite length.

Thus, by the Krull-Schmidt-Remak theorem, we can write $R = I_1 \oplus \cdots \oplus I_n$, with each $I_i \subset R$ an indecomposable module, i.e., an indecomposable ideal. Recall that each $R_i := I_i$ then is a ring with multiplication obtained by restriction from R (and multiplicative identity equal to the projection to $I_i = R_i$ of $1 \in R$). Moreover, we have $R \cong R_1 \times \cdots \times R_n$ (since $R_i \cdot R_j = I_i \cdot I_j \subset I_i \cap I_j = 0$). Each R_i is Artinian, since each I_i is.

Thus, it is enough to show that each R_i is local: this follows since $R_i \cong R_i^{op} \cong \mathrm{End}_{R_i}(R_i) = \mathrm{End}_R(R_i)$ (for the equality $\mathrm{End}_{R_i}(R_i) \cong R_i^{op}$, see Exercise 18.14 below), which is a local ring by Lemma 18.8. □

Exercise 18.14. (Very important) Let R be a (not necessarily commutative) ring.

(i) Show that $\mathrm{End}_R(R) \cong R^{op}$. More precisely (make sure you notice that what follows is a better description, as it gives more information), note that the set-theoretic map $R \to End_R(R)$, sending a to *right* multiplication by a, is a ring homomorphism $R^{op} \to End_R(R)$; show that this ring homomorphism is an isomorphism.

(ii) If M is a left R-module and $n \in \mathbb{N}$, describe a ring isomorphism $\mathrm{End}_R(M^{\oplus n}) \cong M_n(\mathrm{End}_R(M))$.

More generally, describe isomorphisms $\mathrm{Hom}_R(M^{\oplus n}, M^{\oplus m}) \cong M_{m\times n}(\mathrm{End}_R(M))$ of abelian groups as $m, n \in \mathbb{N}$ vary, in a manner respecting composition.

Hint: Send T to $[a_{ij}]$, where $a_{ij} : M \to M$ is $M \overset{j\text{-th copy}}{\hookrightarrow} M^{\oplus n} \to M^{\oplus m} \overset{i\text{-th copy}}{\twoheadrightarrow} M$.

(iii) Put the above two exercises together to give, for any $n \in \mathbb{N}$, an explicit isomorphism $\mathrm{End}_R(R^n) \to M_n(R^{op})$.

Note: Thus, this works exactly as for linear transformations of vector spaces over fields – we haven't changed the description of the isomorphism – the only difference is that we now have to distinguish between R and R^{op}, which are the same only when R is commutative.

18.3. **The theorem of Artin and Wedderburn – I.** Recall that we have been writing 'semisimple' for 'left semisimple'. We will show below, as a consequence of the theorem of Artin-Wedderburn, as to why 'left semisimple' is the same as 'right semisimple', so we temporarily start distinguishing between the two again.

The aim of this subsection is to prove the following theorem:

Theorem 18.15 (Artin-Wedderburn)**.** *Let R be a ring.*

(i) R is left semisimple if and only if there exists a direct product decomposition (as rings)

$$R \cong M_{n_1}(D_1) \times \cdots \times M_{n_r}(D_r).$$

(ii) Given an identification $R \cong M_{n_1}(D_1) \times \cdots \times M_{n_r}(D_r)$, there are exactly r pairwise nonisomorphic simple left modules over R, namely the $D_i^{n_i}$, each $D_i^{n_i}$ viewed as an R-module via left-multiplication by the quotient $M_{n_i}(D_i)$ of R.

(iii) Given two decompositions

$$R \cong \prod_{i=1}^{r} M_{n_i}(D_i) = \prod_{i=1}^{s} M_{n_i'}(D_i'),$$

as rings, we have $r = s$, and after a permutation of the $M_{n_i'}(D_i')$ if necessary we can ensure that $D_i \cong D_i'$ and $n_i = n_i'$ for all i.

Corollary 18.16. *R is left semisimple if and only if it is right semisimple.*

Proof, assuming Theorem 18.15. By Theorem 18.15(i) and its obvious analogue for right R-modules, both the left semisimplicty and the right semisimplicity of R are equivalent to R being isomorphic to a product $\prod_{i=1}^{r} M_{n_i}(D_i)$ with each D_i a division ring. □

Now we prove one implication of Theorem 18.15(i) – a proof of one of the two what may hopefully be called 'Artin-Wedderburn realizations' of R:

Proposition 18.17. *We can write the left R-module R as*

$$R = \bigoplus_{i=1}^{r} M_i^{n_i},$$

where $\{M_i\}_{i=1}^r$ is a set of representatives for the isomorphism classes of simple submodules of the left R-module R, so in particular $M_i \ncong M_j$ if $i \neq j$, and where each n_i is a positive integer: note that such a decomposition exists, since R is of semisimple and (being Artinian) of finite length as a left R-module. For any such decomposition we have

$$R \cong \prod_{i=1}^{r} M_{n_i}(D_i),$$

with each D_i the division ring $\mathrm{End}_R(M_i)^{op}$.

Remark 18.18. Though the proof is easy, it seems an important enough theme to separately write down the philosophy.

- To study an object, study whatever that object acts on: study representations to study groups, study modules with R-action to study a ring R. Specifically, if R acts on an S-module M, we get a homomorphism $R \to \mathrm{End}_S(M)$.
- If M is a simple S-module, then $\mathrm{End}_S(M)$ is a division ring, so R maps to a division ring.
- But this cannot usually be arranged; nevertheless, we can make semisimple R act on $M^{\oplus n}$, where M is a simple S-module for a suitable S. This will give us a homomorphism $R \to \mathrm{End}_S(M^{\oplus n}) \cong M_n(\mathrm{End}_S(M))$ (use Exercise 18.14 for this latter isomorphism).
- How do we get such M and S? R itself is an (R, R)-bimodule, from which, using semisimplicity, we can extract M and S as above ($S = R$ or R^{op} depending on the conventions).

In fact, the proof of Proposition 18.17 will go through the introduction of a series of standard themes which are any way important. One of these is:

Notation 18.19. (i) Let M be a semisimple left module over a ring R, and let M_0 be a simple left R-module. The M_0-isotypic component of M is the sum $N \subset M$ of all submodules of M isomorphic to M_0. If $M = N$ we say that M is M_0-isotypic.

(ii) Thus, any semisimple left R-module M can be written as

$$M = \bigoplus_{i} L_i,$$

where each $L_i \subset M$ is an isotypic component of M, i.e., is M_i-isotypic for some simple left R-module M_i. [54] This decomposition is called the isotypic decomposition of M.

Note that this definition can apply in the setting of a more general abelian category.

Exercise 18.20. Let $M = \bigoplus_i L_i$ be the isotypic decomposition of a semisimple left R-module M.

(i) Show that each endomorphism of M preserves L_i for each i.
(ii) Therefore, restriction to the L_i gives us a homomorphism

$$\mathrm{End}_R(M) \to \prod_i \mathrm{End}_R(L_i).$$

Show that this homomorphism is an isomorphism.

Exercise 18.21. Given an identification $M \cong \bigoplus_i M_i^{\oplus n_i}$ as left R-modules (a finite sum), where the M_i are simple and pairwise nonisomorphic, then show that any endomorphism of M restricts to one of each $M_i^{\oplus n_i}$, giving an isomorphism

$$\mathrm{End}_R(M) \cong \prod_i \mathrm{End}_R(M_i^{\oplus n_i}) = \prod_i M_{n_i}(\mathrm{End}_R(M_i)).$$

Hint: Just combine Schur's lemma, Exercise 18.20, and Exercise 18.14(ii).

Proof of Proposition 18.17. Since R is semisimple as a left module over itself, we have the isotypic decomposition of this module

$$R \cong \bigoplus_i M_i^{\oplus n_i}.$$

Since R is semisimple and hence of finite length as a left R-module, this decomposition is finite. Since $\mathrm{End}_R(R) \cong R^{op}$, we have

$$R \cong \mathrm{End}_R(R)^{op} \cong \mathrm{End}_R\Big(\bigoplus_i M_i^{\oplus n_i}\Big)^{op} \overset{\text{Exercise 18.21}}{\cong} \Big(\prod_i M_{n_i}(\mathrm{End}_R(M_i))\Big)^{op} \overset{D_i' := \mathrm{End}_R(M_i)}{=} \prod_i M_{n_i}(D_i')^{op}$$

(we have commuted *op* with the product). D_i' is a division ring since M_i is simple. To finish getting a product decomposition of R, note that $M_n(D_i')^{op} \cong M_n(D_i)$ where $D_i = (D_i')^{op}$: use $A \mapsto {}^tA$. □

Remark 18.22. In the context of the above proof, note that $\{M_i\}_{i=1}^r$ is also a set of representatives for the isomorphism classes of all simple left R-modules: this follows from the fact that every R-module is a quotient of a free R-module.

[54] It is bad notation to write M_i for a simple module whose isomorphism class appears in M rather than for an isotypic component of M. But we continue with this bad notation, because we still prefer to use M_i for the simple modules in the proofs of the results involving Artin-Wedderburn theorem.

Exercise 18.23. One disadvantage of the above proof of Proposition 18.17 is that it 'uses coordinates': while the decomposition $R \cong \bigoplus_i L_i$ of the left R-module R into isotypic components is canonical, realizing each L_i as $M_i^{\oplus n_i}$ is far from canonical. So how do we describe $L_i \cong M_i^{\oplus n_i}$ using M_i and some coordinate free objects? This is what we will look at in this exercise. So for the first few sub-exercises among the following, go to the more general setting of $M = \bigoplus_i L_i$, where each $0 \neq L_i$ is M_i-isotypic for a simple left R-module M_i. Let $D_i := \mathrm{End}_R(M_i)$; it is a division ring, and M_i is also a left module over D_i. Please be especially careful with the following: I didn't get enough time to do this properly.

(i) Note that the left D_i-action on M_i makes $\mathrm{Hom}_R(M_i, M)$ into a *right* D_i-vector space. If $L_i \cong M_i^{n_i}$, show that this makes L_i into an n_i-dimensional right D_i-vector space.

(ii) Consider the evaluation map

$$ev_i : \mathrm{Hom}_R(M_i, M) \times M_i \to M,$$

given by $(\varphi, x) \mapsto \varphi(x)$. Show that ev_i is D_i-middle linear, and thus gives a left R-module homomorphism

$$ev_i : \mathrm{Hom}_R(M_i, M) \otimes_{D_i} M_i \to M,$$

where the left R-module structure on the left-hand side comes from the left R-action on second factor M_i.

(iii) Show that ev_i is an isomorphism from its source onto the M_i-isotypic subspace $L_i \subset M$, so that the isotypic decomposition of M can be given the following 'coordinate-free description': it is an isomorphism

$$\sum_i ev_i : \bigoplus_i \mathrm{Hom}_R(M_i, M) \otimes_{D_i} M \to M$$

of left R-modules, where $\{M_i\}_i$ is a set of representatives for the isomorphism classes of simple left R-modules appearing in M.
Note: Thus, the 'n_i' in $M_i^{\oplus n_i}$ is more invariantly captured in terms of the n_i-dimensional right D_i-vector space $\mathrm{Hom}_R(M_i, M)$.

(iv) How does $\mathrm{End}_R(M_i^{\oplus n_i}) \cong M_{n_i}(\mathrm{End}_R(M_i))$ get captured in this setting? Note that $\mathrm{End}_R(L_i)$, via its action on L_i, acts on $\mathrm{Hom}_R(M_i, L_i) = \mathrm{Hom}_R(M_i, M)$ by endomorphisms which are tautologically D_i-linear. Show that this gives us an isomorphism of rings

$$\mathrm{End}_R(L_i) \cong \mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, M)).$$

(v) Use this to make the proof of Proposition 18.17 coordinate-free: show that in the above setting we have:

$$\mathrm{End}_R(M) \cong \prod_i \mathrm{End}_R(L_i) \cong \prod_{i=1}^{r} \mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, M)),$$

and specializing to the situation where $M = R$, we get

$$R^{op} \cong \prod_{i=1}^{r} \mathrm{End}_R(L_i) \cong \prod_{i=1}^{r} \mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, R)),$$

each factor of which is the algebra of endomorphisms of the n_i-dimensional right D_i-vector space $\mathrm{Hom}_R(M_i, N)$.

Notation 18.24. As motivated by Exercise 18.23, $\mathrm{Hom}_R(M_i, M)$ is called the multiplicity space of M_i in M.

Remark 18.25. Thus, to summarize in English, the above proof of Proposition 18.17 realizes R^{op} as the product of factors that are in bijection with the simple left R-modules, where the factor corresponding to the simple left R-module M_i is the endomorphism algebra $\mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, M)) \cong M_{n_i}(D_i^{op})$ of the multiplicity space $\mathrm{Hom}_R(M_i, R)$ of M_i in R.

Exercise 18.26. Here is another take on Proposition 18.17. Let $\{M_i\}_{i=1}^r$ be the simple left R-modules up to isomorphism: these are finite in number because these are precisely the simple left R-modules that appear in R, up to isomorphism (see Remark 18.22). Again, please be especially careful about this exercise, which I did not get enough time to work out.

(i) For any ring R, consider the forgetful functor $Forget : R\text{-}Mod \rightsquigarrow AbGrp$. Consider the ring $\mathrm{End}(Forget)$ of endomorphisms of $Forget$, i.e., the ring of natural transformations from $Forget$ to itself. Every element of R defines such a natural transformation: $r \in R$ defines the endomorphism of $Forget(M)$ given by multiplication by r. Show that this defines a ring isomorphism

$$R \to End(Forget).$$

Note/Hint: One can do this as in Problem 4, HW 1, but for an alternative approach note that $Forget(M) = \mathrm{Hom}_R(R, M)$, so $Forget$ "=" h_R, and therefore by the Yoneda lemma, at least set-theoretically:

$$End(Forget) = \mathrm{Nat}(h_R, h_R) \stackrel{\text{Yoneda}}{=} h_R(R) = R^{op}$$

(but this doesn't respect any ring structure). I should add that here I am using an abelian group valued version of the Yoneda lemma. But then you will have to go through the proof of the Yoneda lemma and show that this isomorphism is as described above.

(ii) On the other hand, giving an element $T \in End(Forget)$ is the same as giving a self-isomorphism of each abelian group $Forget(M)$, respecting morphisms $M \to N$ of R-modules. By semisimplicity, this element $T \in End(Forget)$ is completely determined by the values it takes on each M_i, which is an endomorphism $T_i : M_i \to M_i$ of the abelian group underlying M_i.

(iii) Show that any T_i as above commutes with $D_i := \mathrm{End}_R(M_i)$, and that, conversely, given $(T_i : M_i \to M_i)_{i=1}^r$ commuting with each D_i, ie., given $(T_i \in \mathrm{End}_{D_i}(M_i))_{i=1}^r$, there exists a unique element of $\mathrm{End}(Forget)$ that on M_i equals T_i.

(iv) Thus, we get an isomorphism of rings:

$$R \to \prod_{i=1}^{r} \mathrm{End}_{D_i}(M_i).$$

(v) Show that each M_i is a finite-dimensional D_i-vector space.
Hint: If not, choosing an infinite strictly increasing chain of left D_i-vector subspaces $0 = W_0 \subsetneq W_1 \subsetneq W_2 \subsetneq \dots$ of M_i, we get a strictly decreasing chain of left ideals $\mathrm{End}_{D_i}(M_i) = I_0 \supsetneq I_1 \supsetneq I_2 \supsetneq \dots$, where $I_j = \{T \in \mathrm{End}_{D_i}(M_i) \mid T(W_j) = 0\}$. This contradicts the fact that $\mathrm{End}_{D_i}(M_i)$, being a quotient of the left semisimple ring R, is left-Artinian.

Remark 18.27. Thus, we have obtained two different proof of part of the Artin-Wedderburn theorem, giving two different what I would like to call 'Artin-Wedderburn isomorphisms'

$$R \to \prod_{i=1}^{r} \mathrm{End}_{D_i}(V_i),$$

where each V_i is a finite dimensional vector space over a division ring D_i, as follows:

- In the "first/main" and probably more standard proof of Proposition 18.17, [55] we really replace R by R^{op}, and the V_i are not simple left R-modules M_i, but are the multiplicity spaces of the simple left R-modules M_i in the left R-module R. The isomorphism involves the obvious action of R^{op} on the multiplicity spaces (see Exercise 18.23(iv)). D_i is $\mathrm{End}_R(M_i)$, and V_i is a right D_i-vector space.
- In the second (attempted) proof, given in Exercise 18.26, the V_i are the simple left R-modules M_i, and the map $R \to \mathrm{End}_{D_i}(M_i)$ is much simpler: it is just the obvious map with which R acts on the left R-module M_i, namely, the 'action map'. D_i is again just $\mathrm{End}_R(M_i)$, but $V_i = M_i$ is now a left D_i-vector space.
- Thus, we seem to have essentially proved an additional result: the dimension of M_i as a left D_i-vector space is the same as the dimension of its multiplicity space $\mathrm{Hom}_R(M_i, M)$ as a right D_i-vector space. Is there a more direct/conceptual proof of this result, without appealing to the Artin-Wedderburn theorem? I feel there should be, but for now I don't see it.

18.4. **The theorem of Artin and Wedderburn – II.** To go ahead, let us prove that $M_n(D)$ is a left semisimple ring for each natural number n and each division ring D.

Exercise 18.28. Let $R = M_n(D)$, and let $V = D^n$, viewed as a left R-module via left multiplication.

(i) Show that V is a simple left R-module.

[55] In fact, I don't remember seeing the second 'proof' anywhere, so be skeptical of my claim here that it is indeed a proof.

(ii) (Easy) Let $R = M_n(D)$. As a left R-module, show that

$$R = \bigoplus_i M_i,$$

where $V \cong M_i \subset R$ is the submodule consisting of matrices vanishing outside column i. Deduce using (i) above that $R = M_n(D)$ is left semisimple. Thus, these M_i are minimal left ideals of $M_n(D)$. Similarly, $M_n(D)$ is right semisimple.

(iii) Conclude that every simple left R-module is isomorphic to V.

(iv) Show that $\mathrm{End}_R(V)$ is isomorphic to D^{op}.
Hint:
- D^{op} acts on $V = D^n$ by right multiplication (the 'op' ensures that this way, V becomes a left module over D^{op}). Clearly, this action commutes with the R-action on V, so we get $D^{op} \hookrightarrow \mathrm{End}_R(V)$.
- To prove that this is surjective, let $e_1, \ldots, e_n$ be the standard basis of D^n, as a right D-vector space. Show that $e_1 D \subset V$ is the subspace annihilated by all the $\mathrm{diag}(0, a_2, \ldots, a_n)$ in $M_n(D)$. This implies that any element of $\mathrm{End}_R(V)$ stabilizes De_1. Conclude from here that any element of $\mathrm{End}_R(V)$ acts on $e_1 D \cong D$ by right multiplication by an element of D.
- Since V is a simple left R-module, we have $V = Re_1$. Using this or otherwise, show that if $T \in \mathrm{End}_R(V)$ sends e_1 to $e_1 x$ with $x \in D$, then T is given by right multiplication by x. From this conclude that $D^{op} \hookrightarrow \mathrm{End}_R(V)$ is surjective.

(v) Show that $\mathrm{End}_R(V^{\oplus k}) \cong M_k(D^{op})$.
Combine (iv) with Exercise 18.14(ii) above.

Before generalizing the above exercise to its coordinate-free version, let us define a simple ring:

Definition 18.29. A ring R is said to be simple if it has no nonzero proper two-sided ideals.

In this definition I follow Professor Nair's notes, and he seems to follow Lam and Morel; not everyone follows it: in Serge Lang's book, a simple ring is one which is semisimple, and has only one isomorphism class of simple left modules (or equivalently, one isomorphism class of simple left ideals – there can be multiple simple left ideals but they all need to be isomorphic).

These definitions are inequivalent: in the definition that we are following, a simple ring need not be semisimple (see the example from HW 9, copied from Professor Nair's notes, of the ring $k[x, \partial]$ of polynomial differential operators in one variable), but for Serge Lang, a simple ring is semisimple by definition.

Exercise 18.30. This exercise contains a coordinate-free version of much of the above exercise (and slightly more). In place of $M_n(D)$, we consider $\mathrm{End}_D(V)$, where V is a finite dimensional left vector space (we replace left with right/D with D^{op}). $\mathrm{End}_D(V)$ acts on V by definition, and this V takes the place of D^n.

(i) Show that V is an irreducible $\mathrm{End}_D(V)$-module.
(ii) Show that every left ideal in $\mathrm{End}_D(V)$ is of the form I_W, where

$$I_W = \{T \in \mathrm{End}_D(V) \mid T(W) = 0\}.$$

Hint:
- If $T, S \in \mathrm{End}_D(V)$ and $\ker S \subset \ker T$, show that $T = AS$ for some $A \in \mathrm{End}_D(V)$: for this, if $v_1, \ldots, v_r \in V$ are such that $Sv_1, \ldots, Sv_r$ is a left D-basis for $S(V)$, try to define $A(Sv_i) = Tv_i$ for each i; why can such an A be extended to an element of $\mathrm{End}_D(V)$?
- If $S, T \in \mathrm{End}_D(V)$, show that there exist $A, B \in \mathrm{End}_D(V)$ such that $\ker(AS + BT) = \ker S \cap \ker T$. Decompose V into four pieces, one of which is $\ker S \cap \ker T$, and use the above point to dictate what AS and BT have to be on those pieces.

(iii) Similarly, show that the right ideals in $\mathrm{End}_D(V)$ are precisely those of the form J_W, where $W \subset V$ is a sub-D-vector space, and $J_W = \{T \in \mathrm{End}_D(V) \mid T(V) \subset W\}$.
(iv) Conclude that $\mathrm{End}_D(V)$ has no proper nonzero two-sided ideal, so that it is simple.
(v) Using (ii), show that $\mathrm{End}_D(V)$ is a direct sum of finitely many of its simple left ideals, so that it is left semisimple.
(vi) If $R = \mathrm{End}_D(V)$, show that $\mathrm{End}_R(V) \cong D$ (this makes Exercise 18.28(iv) coordinate-free).

For proving Theorem 18.15, we will use one more exercise:

Exercise 18.31. If $R = R_1 \times \cdots \times R_n$ is a product of rings, then show the following.

(i) For each simple left R_i-module V_i, viewing V_i as a left R-module via $R \to R_i$ realizes V_i as a simple left R-module. Moreover, any simple left R-module arises this way for a uniquely determined i.
(ii) Generalize this to giving an equivalence of categories $R\text{-}Mod \rightsquigarrow \prod_{i=1}^n R_i\text{-}Mod$. Thus, giving central idempotents in R gives a product decomposition of R-Mod into (typically simpler) categories of the same type.

Now that we have studied the algebras $M_n(D)$, we can prove Theorem 18.15:

Proof of Theorem 18.15. The "$\Rightarrow$" assertion in (i) of the theorem has been proved in Proposition 18.17. For "$\Leftarrow$", we have seen in Exercise 18.28 above that each $M_n(D)$ is left semisimple, and we saw in Lecture 16 that a product of left semisimple rings is left semisimple. This proves (i).

Exercise 18.31 reduces (ii) to the case where $R = M_n(D)$, which is handled by Exercise 18.28(iii).

Now suppose $R = \prod_{i=1}^r M_{n_i}(D_i) = \prod_{i=1}^s M_{n'_i}(D'_i)$. By (ii), $r = s$ is the number of simple left R-modules of R up to isomorphism, say $M_1, \ldots, M_r$. We may assume after a permutation of the $M_{n_i}(D_i)$ and the $M_{n'_i}(D'_i)$ that M_i is isomorphic to the simple left R-module $D_i^{n_i}$

(resp., $(D_i')^{n_i'}$) on which R acts via the composition of $R \to M_{n_i}(D_i)$ (resp., $R \to M_{n_i'}(D_i')$) and left multiplication. Thus, by Exercise 18.28(iii),

$$(D_i')^{op} \cong \mathrm{End}_{M_{n_i'}(D_i')}(M_i) \cong \mathrm{End}_R(M_i) \cong \mathrm{End}_{M_{n_i}(D_i)}(M_i) \cong D_i^{op},$$

so that $D_i \cong D_i'$.

□

19. Lecture 19 – The Jacobson density theorem and consequences

19.1. Some comments on representations of R and those of $M_n(R)$.

19.1.1. *Modules over R vs modules over $M_n(R)$.* Recall that if M is a left R-module, then $M^{\oplus n}$ is a left $M_n(R)$-module: $M_n(R)$ acts on $M^{\oplus n}$ by 'matrix multiplication'.

The following exercise is a copy of tag 074D of the stacks project: https://stacks.math.columbia.edu/tag/074D , where you will see a couple of more details regarding how to prove it.

Exercise 19.1. Let R be a ring and $n \geqslant 1$. Then:

(i) $M \rightsquigarrow M^{\oplus n}$ defines an equivalence of categories $R\text{-}Mod \rightsquigarrow M_n(R)\text{-}Mod$, with a quasi-inverse given by $N \rightsquigarrow e_{11}N$ (see the explanation in Remark 19.2 below).
(ii) Any two-sided ideal of $M_n(R)$ is of the form $M_n(I) = IM_n(R) = M_n(R)I$, where $I \subset R$ is a two-sided ideal.
(iii) The center $Z(M_n(R))$ of $M_n(R)$ is the center $Z(R)$ of $R \subset M_n(R)$, where R is viewed as the subring of $M_n(R)$ consisting of the scalar matrices.

Remark 19.2. Here is an explanation for what "$N \rightsquigarrow e_{11}N$" means in the statement of the above theorem. $e_{11} \in M_n(R)$ stands for the matrix whose $(1,1)$-th entry is 1 and all other entries are 0. $e_{11}N \subset N$ is then not stable under $M_n(R)$, but it is stable under R, which sits inside $M_n(R)$, as scalar matrices. Thus, $e_{11}N$ is indeed a left R-module.

19.1.2. *Realizing abelian categories as module categories, Morita equivalence.* (Optional, but recommended).

In this subsection, my main reference will be the book Categories and Functors by Pareigis. Except, I might miss some condition written somewhere inside the book, and also I don't know if the terminology in the book is what is universally used these days.

We would like to discuss a more general framework that specializes to the equivalence of categories $R\text{-}Mod \rightsquigarrow M_n(R)\text{-}Mod$. For this, let us state a general theorem that allows us to realize certain abelian categories as module categories: unlike the Freyd-Mitchell embedding theorem, this one doesn't cover all small categories, but it seems to be simpler, gives an equivalence of categories instead of just an embedding, and is perhaps more often concretely realizable. We will assume our categories to be locally small.

Let $\mathcal{A}$ be an abelian category, and let $P \in \operatorname{Ob}\mathcal{A}$. We make the following observations:

- $\operatorname{Hom}_{\mathcal{A}}(P, -)$ is a functor $A \rightsquigarrow AbGrp$.
- Further, there is an obvious ring that acts on each $\operatorname{Hom}_{\mathcal{A}}(P, A)$, namely, $\operatorname{Hom}_{\mathcal{A}}(P, P)$ acts on the right of $\operatorname{Hom}_{\mathcal{A}}(P, A)$ by precomposition, so that $\operatorname{Hom}_{\mathcal{A}}(P, A)$ is a left module over $R := \operatorname{Hom}_{\mathcal{A}}(P, P)^{op}$.
- Thus, for any $P \in \operatorname{Ob}\mathcal{A}$, we get a functor $\mathcal{A} \rightsquigarrow R\text{-}Mod$, where $R = \operatorname{Hom}_{\mathcal{A}}(P, P)^{op}$. Thus, the question is: what properties of $\mathcal{A}$ can ensure that this is an equivalence of categories $\mathcal{A} \rightsquigarrow R\text{-}Mod$?

Example: If we take $\mathcal{A} = R\text{-}Mod$ and $P = R$, then $\mathrm{Hom}_{\mathcal{A}}(P, -)$ is the identity functor $R\text{-}Mod \rightsquigarrow R\text{-}Mod$.

- An object $P \in \mathrm{Ob}\,\mathcal{A}$ is called a generator, if for all objects $A, B \in \mathrm{Ob}\,\mathcal{A}$ and all morphisms $f, g : A \to B$ in $\mathcal{A}$ with $f \neq g$, we have a morphism $h : P \to A$ with $f \circ h \neq g \circ h$: in other words, if $\mathrm{Hom}_{\mathcal{A}}(P, -)$ is a faithful functor.
- We will be interested in projective objects $P \in \mathrm{Ob}\,\mathcal{A}$: this is to ensure that $\mathrm{Hom}_{\mathcal{A}}(P, -)$ is exact.
- A projective object $P \in \mathrm{Ob}\,\mathcal{A}$ is called finite if $\mathrm{Hom}_{\mathcal{A}}(P, -)$ preserves small coproducts (which are not necessarily finite): this means that the compositions $\mathrm{Hom}_{\mathcal{A}}(P, A_j) \to \mathrm{Hom}_{\mathcal{A}}(P, \bigoplus_i A_i)$ (as j vary over the relevant indexing set) induce an isomorphism $\bigoplus_i \mathrm{Hom}_{\mathcal{A}}(P, A_i) \to \mathrm{Hom}_{\mathcal{A}}(P, \bigoplus_i A_i)$: in other words, if P is 'small enough' that every morphism from P to $\bigoplus_i A_i$ factors through the sum of a finite subcollection of the A_i.

  After all, note that $\mathrm{Hom}_{R\text{-}Mod}(R, -)$ certainly commutes with small coproducts, so this is indeed something we might like to impose.
- A finite generator in $\mathcal{A}$ is called a progenerator.

Let us study the above properties.

Exercise 19.3. (i) If $\mathcal{A}$ has small coproducts, show that an object $P \in \mathrm{Ob}\,\mathcal{A}$ is a generator if and only if for all $A \in \mathrm{Ob}\,\mathcal{A}$, there is an epimorphism $\coprod P \to A$, from some small coproduct of copies of P to A. This is almost an 'if and only if', if we understand what to replace 'small' with.

Hint: Take a coproduct of copies of P indexed by $\mathrm{Hom}_{\mathcal{A}}(P, A)$.

(ii) Show that a projective left R-module is a generator for $R\text{-}Mod$ if and only if there exists a surjection $P^{\oplus n} \to R$ for some positive integer n, or equivalently, if and only if R is a direct summand of $P^{\oplus n}$ for some n.

Thus, for instance, a lot of projective modules one can think of, including nonzero free left R-modules, are projective generators for $R\text{-}Mod$. However, $\mathbb{Z}/2\mathbb{Z}$, despite being a projective $\mathbb{Z}/6\mathbb{Z}$-module, is clearly not a projective generator for $\mathbb{Z}/6\mathbb{Z}\text{-}Mod$.

(iii) If $\mathcal{A} = R\text{-}Mod$, show that a projective left R-module is a finite object of $\mathcal{A}$ if and only if it is finitely generated.

Hint: To show that a finite projective left R-module P is finitely generated, let $R^J \twoheadrightarrow P$, and split it as $P \to R^J$. Apply the coproduct condition with the components R of R^J.

Recall that the question is: when is $\mathrm{Hom}_{\mathcal{A}}(P, -) : \mathcal{A} \rightsquigarrow R\text{-}Mod$, where $R := \mathrm{Hom}_{\mathcal{A}}(P, P)^{op}$, an equivalence of abelian categories?

We state the following theorem without proof: I have copied it from Theorem 1 in Section 4.11 of the book 'Categories and functors' by Bodo Pareigis, but I could have made some careless error in not paying attention to some condition imposed somewhere in the book:

Theorem 19.4. *Let $\mathcal{A}$ be a locally small abelian category. Then:*

(i) There exists an equivalence of categories $E : \mathcal{A} \rightsquigarrow R\text{-}Mod$, for some ring R, if and only if $\mathcal{A}$ contains a progenerator $P \in \operatorname{Ob}\mathcal{A}$ such that arbitrary small coproducts of copies of P exist in $\mathcal{A}$.

(ii) Given any equivalence $E : \mathcal{A} \rightsquigarrow R\text{-}Mod$, we may choose the progenerator P such that $R = \operatorname{Hom}_{\mathcal{A}}(P,P)^{op}$, and such that E is naturally isomorphic to $\operatorname{Hom}_{\mathcal{A}}(P,-) : \mathcal{A} \rightsquigarrow R\text{-}Mod$.

Here are some very brief and impressionistic comments on the proof of the bit that is of interest to us, that if $P \in \operatorname{Ob}\mathcal{A}$ is a progenerator and arbitrary small coproducts of P exist in $\mathcal{A}$, then $\operatorname{Hom}_{\mathcal{A}}(P,-)$ defines an equivalence of categories $\mathcal{A} \to R\text{-}Mod$, where $R = \operatorname{Hom}_{\mathcal{A}}(P,P)^{op}$. In particular, we want to show that for $X, Y \in \operatorname{Ob}\mathcal{A}$, $\operatorname{Hom}_{\mathcal{A}}(P,-)$ induces a bijection

$$\operatorname{Hom}_{\mathcal{A}}(X,Y) \to \operatorname{Hom}_R(\operatorname{Hom}_{\mathcal{A}}(P,X), \operatorname{Hom}_{\mathcal{A}}(P,Y)).$$

The injectivity is anyway clear from P being a generator (faithfulness). For $X = Y = P$, we do get an isomorphism

$$\operatorname{Hom}_{\mathcal{A}}(P,P) \to \operatorname{Hom}_R(\operatorname{Hom}_{\mathcal{A}}(P,P), \operatorname{Hom}_{\mathcal{A}}(P,P))$$

– this is just a consequence of the identification $\operatorname{End}_R(R) \cong R^{op}$. From this, we can make this work in the case where X and Y are small coproducts of copies of P. For the general case, write $X, Y \in \operatorname{Ob}\mathcal{A}$ as cokernels of objects of the form $\bigoplus_i P$ and $\bigoplus_j P$, and apply the given conditions on P. Note that by what we have seen about projective resolutions, morphisms $X \to Y$ can be lifted to morphisms between coproducts of copies of P that surject to them. This ends the informal sketch of some of the ideas used in the proof of the above theorem. For a more detailed but succinct proof, you may also see the following notes: https://math.huji.ac.il/∼ayomdin/Notes/Ma120c.pdf

In any case, notice a corollary (use Exercise 19.3 as well):

Corollary 19.5 (Morita)**.** *The categories $R\text{-}Mod$ and $S\text{-}Mod$ are equivalent if and only if there exists a finitely generated projective generator P for $R\text{-}Mod$ such that $S \cong \operatorname{End}_R(P)^{op}$.*

Definition 19.6. Rings R and S are said to be Morita equivalent if $R\text{-}Mod$ and $S\text{-}Mod$ are equivalent.

Example 19.7. It is immediate that $P := R^n$ is a finitely generated projective generator for $R\text{-}Mod$, so $M \rightsquigarrow \operatorname{Hom}_R(R^n, M) \cong M^{\oplus n}$, with each $M^{\oplus n}$ viewed as a left module over $(\operatorname{End}_{R\text{-}Mod}(R^n))^{op} \cong M_n(R)$, gives us an equivalence of categories $R\text{-}Mod \rightsquigarrow M_n(R)\text{-}Mod$.

The equivalence of categories $R\text{-}Mod \rightsquigarrow S\text{-}Mod$ given by Corollary 19.5 can be made more explicit. For this, note the following:

- In Corollary 19.5, the fact that $S \cong \operatorname{End}_R(P)^{op}$ also makes P into a right S-module, and hence in fact an (R,S)-bimodule (the R-action and the S-action clearly commute).

- Thus, the S-module structure on $\mathrm{Hom}_R(P, A)$ is just the S-module structure it gets from the R-S-bimodule structure on P.

These considerations with some work let us deduce from Theorem 19.4 the following, which will also be stated without proof (and is copied from Pareigis' book as well):

Theorem 19.8 (Morita)**.** *Let R, S be rings, and let P be an (R, S)-bimodule. Then the following are equivalent:*

(i) $P \otimes_S - : S\text{-}Mod \rightsquigarrow R\text{-}Mod$ *is an equivalence of categories.*
(ii) $- \otimes_R P : Mod\text{-}R \rightsquigarrow Mod\text{-}S$ *is an equivalence of categories.*
(iii) $\mathrm{Hom}_R(P, -) : R\text{-}Mod \rightsquigarrow S\text{-}Mod$ *is an equivalence of categories.*
(iv) $\mathrm{Hom}_S(P, -) : Mod\text{-}S \rightsquigarrow Mod\text{-}R$ *is an equivalence of categories.*
(v) As a left R-module, P is a progenerator for R-Mod, and the right action of S on P defines an isomorphism $S \cong \mathrm{End}_R(P)^{op}$.
(vi) As a right S-module, P is a progenerator for Mod-S, and the left action of R on P defines an isomorphism $R \cong \mathrm{End}_S(P)$.

Moreover, such (R, S)-bimodules exist when R and S are Morita equivalent.

19.2. **The setting for the density theorems.** Recall that giving a left R-module structure on M is the same as giving a ring homomorphism $R \to \mathrm{End}_{\mathbb{Z}}(M)$. Then $\mathrm{End}_R(M) \subset \mathrm{End}_{\mathbb{Z}}(M)$ is precisely the centralizer of R (i.e., of the image of R in $\mathrm{End}_{\mathbb{Z}}(M)$).

Let $R' := \mathrm{End}_R(M) \subset \mathrm{End}_{\mathbb{Z}}(M)$. Since $R' \hookrightarrow \mathrm{End}_{\mathbb{Z}}(M)$, M has a left R'-module structure as well: thus, M has the structure of a left R-module and a left R'-module, and these module structures commute: $r \cdot (r' \cdot m) = r' \cdot (r \cdot m)$ for all $r \in R$ and $r' \in R'$. Indeed, this commutativity follows from the fact that the images of R and R' in $\mathrm{End}_{\mathbb{Z}}(M)$ commute.

Remark 19.9.
(i) Instead, since a left R'-module is equivalent to a right $(R')^{op}$-module, we could also view M as an $(R, (R')^{op})$-bimodule: this is the notation used in Professor Nair's notes, and in various other sources such as Morel's notes. However, we will follow the convention of Serge Lang, to just treat M as having two commuting left R-module structures.
(ii) Notice that we have seen this consideration already in the material of the previous subsection on Morita equivalence: there the left R-module P was viewed as an (R, S)-bimodule, where $S = \mathrm{End}_R(P)^{op}$. We also made a similar consideration in Lecture 18, in an exercise that attempted to give an alternate take on the proof of the Artin-Wedderburn theorem.

Coming back to the left R-module M and $R' = \mathrm{End}_R(M) \subset \mathrm{End}_{\mathbb{Z}}(M)$, we claim that M has an obvious structure of an $R \otimes_{\mathbb{Z}} R'$-module (we will not need this, but it seems to me to help conceptual clarity).

Exercise 19.10. Let S_1, S_2 be R-algebras, where R is a commutative ring. Define $\iota_1 : S_1 \to S_1 \otimes_R S_2$ and $\iota_2 : S_2 \to S_1 \otimes_R S_2$ by $\iota_1(s_1) = s_1 \otimes 1$ and $\iota_2(s_2) = 1 \otimes s_2$. Then, for

any R-algebra S, $(- \circ \iota_1, - \circ \iota_2)$ induces a bijection:

$$\begin{array}{c} \mathrm{Hom}_{R\text{-}Alg}(S_1 \otimes_R S_2, S) \overset{(-\circ\iota_1, -\circ\iota_2)}{\to} \\ \{(f_1, f_2) \in \mathrm{Hom}_{R\text{-}Alg}(S_1, S) \times \mathrm{Hom}_{R\text{-}Alg}(S_2, S) \mid f_1(S_1), f_2(S_2) \subset S \text{ commute}\}. \end{array}$$

Hint: This exercise generalizes the assertion in Lecture 8 that tensor product over R is a coproduct in the category of commutative R-algebras. The same proof goes through to prove this more general assertion, with only the requirement that $f_1(S_1)$ and $f_2(S_2)$ commute.

Thus, by Exercise 19.10, one in fact has a left $R \otimes_{\mathbb{Z}} R'$-module structure on M.

Since any left R-module M gives rise to a left R'-module structure, we can repeat this, and form

$$R'' := \mathrm{End}_{R'}(M) = \text{ the centralizer of } R' \text{ in } \mathrm{End}_{\mathbb{Z}}(M) \subset \mathrm{End}_{\mathbb{Z}}(M).$$

Just as the left R-module M can be viewed as a left R'-module, the left R'-module M can also be viewed as a left R''-module. Moreover, since $\mathrm{Image}(R \to \mathrm{End}_{\mathbb{Z}}(M))$ commutes with R', $\mathrm{Image}(R \to \mathrm{End}_{\mathbb{Z}}(M))$ is contained in $R'' \subset \mathrm{End}_{\mathbb{Z}}(M)$, giving us a ring homomorphism

$$\text{'action map'} : R \to R'' \subset \mathrm{End}_{\mathbb{Z}}(M).$$

We call this the action map since it is just R mapping to $\mathrm{End}_{\mathbb{Z}}(M)$ by its obvious action on M.

In density theorems, we are interested in the following question: Let M be a left R-module, form the centralizer R' of (the image of) R in $\mathrm{End}_{\mathbb{Z}}(M)$, and then the centralizer $R'' = \mathrm{End}_{R'}(M)$. When is $R \to R''$ an isomorphism, or at least a surjection? Theorems that give conditions under which $R \to R''$ is an isomorphism or a surjection are called double centralizer theorems.

19.3. **Jacobson density theorem.**

Theorem 19.11 (Jacobson density theorem: Jacobson, Chevalley)**.** *Let M be a semisimple left R-module, and let $R' = \mathrm{End}_R(M)$ and $R'' = \mathrm{End}_{R'}(M)$ as above. Consider the morphism $R \to R''$.*

(i) The image of $R \to R''$ is dense in the following sense: for all $x_1, \ldots, x_n \in M$, $\exists\, a \in R$ such that $a \cdot x_i = u \cdot x_i$, $\forall\, 1 \leqslant i \leqslant n$. [56]

(ii) If M is finitely generated over R', then $R \to R''$ is surjective (and hence also injective if M was a faithful left R-module).

[56] Thus, 'density' in the sense that the action of an element of R'' on any finite collection of elements can be captured by an element of R on those elements. I am unaware of/haven't looked up topological formalizations of this notation of density.

Remark 19.12. Before proving the theorem, let me ramble about my attempts to make sense of the theorem, [57] in the special case where we put the following additional assumptions on M: M is of finite length over R, and writing $M \cong \bigoplus_i M_i^{n_i}$ with each M_i simple, each M_i is finite dimensional over $D_i := \mathrm{End}_R(M_i)$. Be especially skeptical of the following.

(i) Since M is semisimple, we have an isotypic decomposition:

$$M \cong \bigoplus_i M_i^{n_i} \cong \bigoplus_i \mathrm{Hom}_R(M_i, M) \otimes_{D_i} M_i$$

from an exercise in the notes for Lecture 18, where $D_i = \mathrm{End}_R(M_i)$ is a division ring, and the n_i-dimensional right D_i-vector space $\mathrm{Hom}_R(M_i, M)$ is the multiplicity space for M_i in M. Here, $\mathrm{Hom}_R(M_i, M) \otimes_{D_i} M_i$ maps to M via the evaluation map that sends each $\varphi_i \otimes x_i$ to $\varphi_i(x_i)$.

(ii) By an exercise from Lecture 18, then

$$R' = \mathrm{End}_R(M) \cong \prod_i \mathrm{End}_R(M_i^{\oplus n_i}) \cong \prod_i \mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, M)).$$

Now when we view $M \cong \bigoplus_i \mathrm{Hom}_R(M_i, M) \otimes_{D_i} M_i$ as a left R'-module, each $\mathrm{Hom}_R(M_i, M)$ is a simple left module over $\mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, M))$, and now the M_i take on the role of the multiplicity spaces. We have assumed $\dim_{D_i} M_i < \infty$.

(iii) Therefore, by the same exercise, we should have $R'' = \prod_i \mathrm{End}_{D_i}(M_i)$. Thus, it seems that the Jacobson density theorem is basically giving the surjectivity of the action map $R \to \prod_i \mathrm{End}_{D_i}(M_i)$, whenever there are only finitely many M_i involved and each occurs with finite multiplicity. Anyway, we will see this theme in some of the corollaries to the Jacobson density theorem below.

Note that this is also what our other tentative second take on the Artin-Wedderburn theorem, from an exercise in Lecture 18, sought to prove, correctly or otherwise.

(iv) Anyway, what I wish to emphasize is the following relationship between how R' and R'' act on M (and R – after all, R itself acts through R''): $R' = \prod_i \mathrm{End}_{D_i}(\mathrm{Hom}_R(M_i, M))$ and $R'' = \prod_i \mathrm{End}_{D_i}(M_i)$ are products of simple rings indexed by the same set and involving the same division algebras D_i, and the set up gives us a natural *bijection*

$$\{\text{Simple left submodules of } R' \text{ in } M\} \to \{\text{Simple left submodules of } R'' \text{ in } M\},$$

say $\sigma_i \mapsto \tau_i$, such that as a representation of $R' \otimes_{\mathbb{Z}} R''$ we have a decomposition

$$M = \bigoplus_i \sigma_i \otimes \tau_i,$$

where $\sigma_i \otimes \tau_i$ is viewed as a left module over $R' \otimes_{\mathbb{Z}} R''$ using the universal property of the tensor product of algebras discussed in Exercise 19.10 above. This seems reminiscent of many 'duality' type theories in the literature, such as the Schur-Weyl duality and the Howe duality. So there could be a better way to state this theorem, but I could be making mistakes here.

Now we move onto the usual proof of the theorem.

[57] This was not discussed in the lecture, and is very optional.

Remark 19.13. (i) Notice the difference in emphasis from the approach to the Artin-Wedderburn theorem: in proving it we made a fixed semisimple ring R act on various modules, whereas here only the module M is semisimple (and not R), and we are making different rings act on M.

(ii) To repeat, this is a theorem of the 'double centralizer' genre.

Let us rephrase the assertion of (i) of the theorem:

- Rephrasing: It says that for all $u \in R''$ and $(x_1, \dots, x_n) \in M^{\oplus n}$, there exists $a \in R$ such that $u \cdot (x_1, \dots, x_n) = a \cdot (x_1, \dots, x_n)$, where

$$u \cdot (x_1, \dots, x_n) = (u \cdot x_1, \dots, u \cdot x_n), \qquad \text{and } a \cdot (x_1, \dots, x_n) = (a \cdot x_1, \dots, a \cdot x_n). \tag{75}$$

- Another rephrasing: Make R and R'' act diagonally on $M^{\oplus n}$ (i.e., as in (75)). Then every cyclic R-submodule of $M^{\oplus n}$ is stable under R''.

Thus, the $n = 1$ case of (i) of the theorem is a special case of the following lemma:

Lemma 19.14. *Let M be a semisimple left R-module, and let $R' = \mathrm{End}_R(M)$ and $R'' = \mathrm{End}_{R'}(M)$ as above. Then*

$$\{R\text{-submodules of } M\} = \{R''\text{-submodules of } M\}.$$

In other words, any R-submodule $N \subset M$ is stabilized by R''.

Proof. Since M is semisimple, we can write $M = N \oplus N'$ as left R-modules. We need to show that any $u \in R''$ stabilizes N. Write $\mathrm{pr}_N : M \to N$ for the unique projection onto N with N' as its kernel. Then clearly $\mathrm{pr}_N \in R'$, so u commutes with pr_N: $u \circ \mathrm{pr}_N = \mathrm{pr}_N \circ u$, so that $u(N) = u \circ \mathrm{pr}_N(M) = \mathrm{pr}_N \circ u(M) \subset N$, as desired. □

Remark 19.15. Thus, the point of the proof seems to be as follows: we want to say that R'' is not too large, for which we want R' to be large enough. Since M is semisimple, we had lots of projections pr_N, which gave us enough elements in R' for it to be large enough, and thus cut R'' down.

Now we can prove the general case of Theorem 19.11:

Proof of Theorem 19.11. First assume (i), and let us show (ii). If $M = R'x_1 + \cdots + R'x_n$, then any $u \in R'' = \mathrm{End}_{R'}(M)$ is completely determined by $u \cdot x_1, \dots, u \cdot x_n$. Thus, if we get $a \in R$ such that $u \cdot x_i = a \cdot x_i$ for each i, then u and the image of a under $R \to R''$ would agree on each x_i, and hence be equal as elements of R'', giving (ii).

Now it suffices to prove (i). As in (75), we make R and R'' act 'diagonally' on $M^{\oplus n}$. All we need to show is that $R \cdot (x_1, \dots, x_n) \subset M^{\oplus n}$ is stable under R''. Since we have treated the $n = 1$ case, this will follow if we show that the action of R'' on $M^{\oplus n}$, or rather the image of $R'' \to \mathrm{End}_{\mathbb{Z}}(M^{\oplus n})$, is contained in the analogue of R'' with M replaced by $M^{\oplus n}$: in other words, it is enough to show that the image of $R'' \to \mathrm{End}_{\mathbb{Z}}(M^{\oplus n})$ commutes with $\mathrm{End}_R(M^{\oplus n})$ inside the big ring $\mathrm{End}_{\mathbb{Z}}(M^{\oplus n})$.

We have, by an exercise from Lecture 18, an identification $\mathrm{End}_{\mathbb{Z}}(M^{\oplus n}) = M_n(\mathrm{End}_{\mathbb{Z}}(M))$. Under this identification:

- $\mathrm{End}_R(M^{\oplus n})$ identifies with $M_n(\mathrm{End}_R(M)) = M_n(R') \subset M_n(\mathrm{End}_{\mathbb{Z}}(M))$; and
- The action of $u \in R'' \subset \mathrm{End}_{\mathbb{Z}}(M)$ identifies with $\mathrm{diag}(u, \cdots, u) \in M_n(\mathrm{End}_{\mathbb{Z}}(M))$ (because R'' acts 'diagonally' on $M^{\oplus n}$).

Thus, it is enough to show that $M_n(R')$ commutes with $\mathrm{diag}(u, \cdots, u)$ in the big ring $M_n(\mathrm{End}_{\mathbb{Z}}(M))$, which follows from the fact that $u \in R'' \subset \mathrm{End}_{\mathbb{Z}}(M)$ commutes with $R' \subset \mathrm{End}_{\mathbb{Z}}(M)$. □

19.4. **Some examples related to the Jacobson density theorem.**

Example 19.16. Only the first of the following examples is an application of the Jacobson density theorem (that too, only to recover something we already know). The rest are just illustrations of it being satisfied in relatively more 'real world' situations; they are meant to give an idea of why you should expect to see some double centralizers.

(i) If V is a finite dimensional vector space over a division ring $R = D$, then by definition, $R' = \mathrm{End}_D(V)$ is the centralizer of $D \subset \mathrm{End}_{\mathbb{Z}}(V)$. V is a simple and hence finitely generated left $\mathrm{End}_D(V)$-module (this simplicity was an exercise in Lecture 18, and is easy to see). Thus, by the Jacobson density theorem, the centralizer R'' of R' in V is D, something we had seen in an exercise in Lecture 18.

(ii) Let V, W be finite dimensional vector spaces over a field k. A problem from HW 3 asked you to prove that the obvious map $\mathrm{End}_k(V) \otimes_k \mathrm{End}_k(W) \to \mathrm{End}_k(V \otimes W)$, defined simply by the functoriality of the tensor product (i.e., it sends $T \otimes S$, where $T : V \to V$ and $S : W \to W$, to $T \otimes S : V \otimes_k W \to V \otimes_k W$), is an isomorphism of vector spaces. Clearly, it respects composition and is hence an isomorphism of k-algebras. One way to see that it is an isomorphism of vector spaces is by showing that the following diagram is commutative:

$$\begin{array}{ccccc} \mathrm{End}_k(V) \otimes_k \mathrm{End}_k(W) & & \longrightarrow & & \mathrm{End}_k(V \otimes_k W) \\ \cong \downarrow & & & & \cong \uparrow \\ (V^\vee \otimes_k V) \otimes_k (W^\vee \otimes_k W) & \xrightarrow{\cong} & (V^\vee \otimes_k W^\vee) \otimes_k (V \otimes_k W) & \xrightarrow{\cong} & (V \otimes_k W)^\vee \otimes_k (V \otimes_k W) \end{array}.$$

This commutativity can be checked, e.g., using basis elements.

In any case, since $\mathrm{End}_k(V) \otimes_k \mathrm{End}_k(W) \to \mathrm{End}_k(V \otimes W)$ is an isomorphism of k-algebras, we can consider $\mathrm{End}_k(V) \subset \mathrm{End}_k(V \otimes_k W)$ and $\mathrm{End}_k(W) \subset \mathrm{End}_k(V \otimes_k W)$: the former inclusion is given by $T \mapsto T \otimes \mathrm{id}_W$, and the latter by $S \mapsto \mathrm{id}_V \otimes S$.[58]

Show as an exercise that inside $\mathrm{End}_k(V \otimes_k W)$, $\mathrm{End}_k(V)$ and $\mathrm{End}_k(W)$ are centralizers of each other. I don't see that this is proved by the Jacobson density

[58] In contrast, note that $\mathrm{End}_k(V \oplus W)$ does not have such a nice description in terms of $\mathrm{End}_k(V)$ or $\mathrm{End}_k(W)$.

theorem, but it is certainly consistent with it: the module $V \otimes_k W$, being finite dimensional over k, is finitely generated over either of $\mathrm{End}_k(V)$ and $\mathrm{End}_k(W)$, and is also semsimple over these, because $\mathrm{End}_k(V)$ and $\mathrm{End}_k(W)$ are semisimple rings.

(iii) Let R be a semisimple ring. Think of R as a left R-module, via left multiplication. Via right multiplication, R is also a right R^{op}-module. It is clear that in $\mathrm{End}_{\mathbb{Z}}(R)$, R and R^{op} are the centralizers of each other. Thus, the Jacobson density theorem is satisfied in this situation.

Here is a slightly more concrete example where one sees this in action. Let D be a finite dimensional division algebra over k: thus, D is a k-algebra which is a division ring, and $\dim_k D < \infty$. Let $A = M_n(D)$: then A is also a finite dimensional k-algebra; it is a simple ring (as we saw in Lecture 18), but not a division ring. We have a ring homomorphism

$$A \to \mathrm{End}_k(A) \qquad \text{and} \qquad A^{op} \to \mathrm{End}_k(A),$$

the former sending $a \in A$ to the left multiplication $l_a : A \to A$, and the latter sending $a \in A^{op} = A$ to the right multiplication $r_a : A \to A$.

The images of A and A^{op} in $\mathrm{End}_k(A)$ clearly commute with each other (as left and right multiplications commute), so by the universal property for tensor products of algebras, we get a homomorphism

$$A \otimes_k A^{op} \to \mathrm{End}_k(A).$$

We will quote later (see Lemma 19.37 below) that $A \otimes_k A^{op}$ is a simple k-algebra, so the above map is injective. Thus, by dimension considerations, it is surjective (both sides have dimension $(\dim_k A)^2$). Thus, not only are $A, A^{op} \subset \mathrm{End}_k(A)$ centralizers of each other, they together complementarily constitute $\mathrm{End}_k(A)$ in some sense.

19.5. **Some corollaries of the Jacobson density theorem.** The following corollary is immediate from the Jacobson density theorem, and describes the images of rings/group rings under irreducible representations:

Corollary 19.17. *(i) Let M be a simple left R-module, so that $D := R' = \mathrm{End}_R(M)$ is a division ring. Assume that M is of finite dimension n as a left vector space over D. Then*

$$R \to R'' = \mathrm{End}_{R'}(M) = \mathrm{End}_D(M) \cong M_n(D^{op})$$

is surjective.

(ii) Suppose R is a (possibly infinite dimensional) algebra over a field k, and let M be a simple left R-module such that $\dim_k M < \infty$; [59] *note that $D := R' = \mathrm{End}_R(M)$ is a division ring. Then*

$$R \to R'' = \mathrm{End}_{R'}(M) = \mathrm{End}_D(M) \cong M_n(D^{op})$$

is surjective, where $n = \dim_D M$ is (being asserted to be) finite.

[59] This is automatic if $\dim_k R < \infty$, since in this case, for any $x \in M$, $r \mapsto rx$ defines a surjection $R \to M$).

(iii) *If G is a (possibly infinite) group and (ρ, V) is an irreducible finite dimensional representation of G over a field k, then the map $\rho : k[G] \to \mathrm{End}_k(V)$ (extended k-linearly from $\rho : G \to GL_k(V)$) has image equal to $\mathrm{End}_D(V) \subset \mathrm{End}_k(V)$, where D is the finite dimensional division k-algebra $\mathrm{End}_{k[G]}(V) =: \mathrm{End}_G(V)$.*

(iv) *(Burnside's theorem) In the situation of (ii), assume that $k = \bar{k}$ is algebraically closed. Then D equals k, so that the obvious map $R \to \mathrm{End}_k(M)$ is surjective, i.e., we have $R \twoheadrightarrow R'' = \mathrm{End}_k(M) = M_n(k^{op}) = M_n(k)$.*

(v) *In the situation of (iii), assume further that $k = \bar{k}$ is algebraically closed. Then $\rho : k[G] \to \mathrm{End}_k(V)$ is surjective, giving a surjection $k[G] \to \mathrm{End}_k(V) \cong M_n(k)$ for some n.*

Proof. (i) is an immediate consequence of the Jacobson density theorem, Theorem 19.11(ii).

(ii) follows from (i), since $\dim_D M \leqslant \dim_k M < \infty$.

(iii) is a special case of (ii), where $R = k[G]$: note that $R \to R'' \subset \mathrm{End}_{\mathbb{Z}}(V)$, being the 'action map', is just $\rho : k[G] \to \mathrm{End}_k(V)$.

To conclude (iv) from (ii), since the D of (iii) is a finite dimensional k-algebra, it suffices to show the following: if D is a division algebra over a field k such that $\dim_k D < \infty$, and if k is algebraically closed, then $k = D$. This follows from Lemma 19.18 below.

Finally, (v) is a special case of (iv), where $R = k[G]$. □

Lemma 19.18. *If k is an algebraically closed field and D is a finite dimensional division algebra over k, then $D = k$.*

Proof. Let $\alpha \in D \backslash k$, and we will get a contradiction. Since k is in the center of D, $k[\alpha] \subset D$ is a commutative integral domain that is finite dimensional as a k-vector space, and hence isomorphic to $k[x]/(f)$, where $f \in k[x]$ is an irreducible polynomial. But this forces $k[\alpha]$ to be a finite field extension of k, which is nontrivial since $\alpha \notin k$, contradicting the fact that k is algebraically closed. □

Remark 19.19.

(i) In Corollary 19.17(i), the hypothesis that M is of finite dimension as a left vector space over D is not automatic. Indeed, let $M = V$ be an infinite dimensional vector space over a field k, viewed as a module over $R = \mathrm{End}_k(V)$. It is easy to see that R acts transitively on $V \backslash \{0\}$, so that V is a simple R-module, and that $D := R' := \mathrm{End}_R(V)$ equals k. But V is not finite dimensional as a left vector space over k.

(ii) At the risk of repetition, the hypothesis of (i) of Corollary 19.17, namely, that $n = \dim_D M < \infty$, might seem difficult to ensure, but (iii) gives an important and commonly seen situation where that hypothesis is automatic.

Corollary 19.17 deals with simple modules; it has an easy generalization that applies to semisimple modules which satisfy a 'multiplicity one condition' for its simple submodules. To not make our zoo too big, we avoid stating the generalizations of (iv) and (v) of the corollary.

Corollary 19.20. *(i) Let $M = \bigoplus_{i=1}^r M_i$ be a semisimple left R-module, where each M_i is a simple left R-module and $M_i \ncong M_j$ for all $i \neq j$, so that $D_i := \mathrm{End}_R(M_i)$ is a division ring for each i. Assume that each M_i is of finite dimension n_i as a vector space over D_i. Then the 'action map'*

$$R \to R'' = \mathrm{End}_{R'}(M) = \prod_{i=1}^r \mathrm{End}_{D_i}(M_i) \cong \prod_{i=1}^r M_{n_i}(D_i^{op})$$

is surjective.

(ii) Suppose R is a (possibly infinite dimensional) algebra over a field k. Let $M_1, \ldots, M_r$ be simple left R-modules, such that $M_i \ncong M_j$ for all $i \neq j$, and such that $\dim_k M_i < \infty$ for each i. Note that $D_i := \mathrm{End}_R(M_i)$ is a division k-algebra for each i. Then $\dim_k D_i < \infty$ for each i, and the 'action map'

$$R \to \prod_{i=1}^r \mathrm{End}_{D_i}(M_i)$$

is surjective.

(iii) If G is a (possibly infinite) group and $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ are pairwise non-isomorphic irreducible finite dimensional representations of G over k, then the map

$$\rho = \prod_{i=1}^r \rho_i : k[G] \to \prod_{i=1}^r \mathrm{End}_{D_i}(V_i)$$

is surjective, where D_i is the finite dimensional division k-algebra $\mathrm{End}_{k[G]}(V_i)$.

(iv) In the situation of (ii), for each $1 \leqslant i \leqslant r$ there exists $e_i \in R$ which acts as the identity on M_i, and which annihilates M_j for $i \neq j$.

Proof. Let us prove (i), by appropriately generalizing the proof of (i) of Corollary 19.17. Let $M = \bigoplus_{i=1}^r M_i$. By an exercise from Lecture 18, each element of $\mathrm{End}_R(M)$ preserves M_i for each i, and gives an isomorphism

$$R' := \mathrm{End}_R(M) \cong \prod_{i=1}^R \mathrm{End}_R(M_i) = \prod_{i=1}^r D_i$$

(this uses that $M_i \ncong M_j$ for $i \neq j$). Since each M_i is assumed to be a finite dimensional D_i-vector space, it is easy to see that M is a finitely generated R'-module. Therefore, Jacobson's density theorem applies, gives us the surjectivity of the action map

$$R \to R'' = \mathrm{End}_{R'}(M) = \prod_{i=1}^r \mathrm{End}_{D_i}(M),$$

giving (i).

The proofs of (ii) and (iii) are analogous generalizations of the corresponding assertions in Corollary 19.17: one again considers $M := \bigoplus_{i=1}^r M_i$ for the former, and $(\rho, V) := \bigoplus_{i=1}^r (\rho_i, V_i)$ for the latter.

(iv) is an immediate consequence of (ii). □

We get the consequence that the isomorphism class of an irreducible or even semisimple finite dimensional representation of an arbitrary group, in characteristic zero, is determined by its character:

Corollary 19.21. *Let R be a (possibly infinite dimensional) k-algebra, where k is a field of characteristic zero. Let M and N be semisimple left R-modules that are finite dimensional as vector spaces over k. Suppose that for all $a \in R$,*

$$\mathrm{tr}(a|_M) = \mathrm{tr}(a|_N),$$

where the left-hand side denotes the trace of $(m \mapsto am) \in \mathrm{End}_k(M)$, and the right-hand side is similar. Then M and N are isomorphic as left R-modules.

Proof. There exists a finite set $L_1, \ldots, L_r$ of simple left R-modules, and nonnegative integers $p_1, \ldots, p_r$ and $q_1, \ldots, q_r$, such that

$$M \cong \bigoplus_{i=1}^{r} L_i^{\oplus p_i}, \qquad \text{and } N \cong \bigoplus_{i=1}^{r} L_i^{\oplus q_i}$$

(note that we allow some of the p_i and the q_i to be 0, so we are not assuming that M and N contain isomorphic simple submodules). Note that $\dim_k L_i < \infty$ for each i.

It is enough to show that for each $1 \leqslant i \leqslant r$, we have $p_i = q_i$. Use Corollary 19.20(iv) to choose $e_i \in R$ such that e_i acts as the identity on M_i (and hence on L_i), and as 0 on M_j (and hence on L_j) for $j \neq i$. Then, in k we have

$$p_i \cdot (\dim_k L_i) = \mathrm{tr}(e_i|_M) = \mathrm{tr}(e_i|_N) = q_i \cdot (\dim_k L_i).$$

Since $\dim_k L_i \neq 0$ in k [60] we have $p_i = q_i$, as required. □

Remark 19.22. In particular, Corollary 19.21, applied to the group algebra $R = k[G]$ of a possibly infinite group G, where k is a field of characteristic zero, shows that any semisimple finite dimensional representation ρ of G is determined by its character, namely the function $G \ni g \mapsto \mathrm{tr}\,\rho(g) \in k$.

Corollary 19.23. *Let G be a finite group, and k a field, such that $k[G]$ is semisimple.* [61] *Then there are only finitely many irreducible representations $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ of G up to isomorphism, and these representations $\rho_i : G \to GL_k(V_i)$, extended by k-linearity to $k[G] \to \mathrm{End}_k(V_i)$, define an isomorphism*

$$k[G] \overset{\prod_{1 \leqslant i \leqslant r} \rho_i}{\to} \mathrm{End}_{D_i}(V_i) \cong \prod_{i=1}^{r} \mathrm{M}_{n_i}(D_i^{op}),$$

where D_i is the division algebra $\mathrm{End}_{k[G]}(V_i)$, and $n_i = \dim_{D_i} V_i$ $(< \dim_k V_i < \infty)$.

[60] This is because k has characteristic 0.

[61] In Lecture 20, hopefully we will show that $k[G]$ is semisimple if and only if $(\mathrm{char}\, k, \#G) = 1$.

Proof. We already know that a semisimple ring has only finitely many simple left modules up to isomorphism, so we let them be $(\rho_1, V_1), \ldots, (\rho_r, V_r)$. By Corollary 19.20(iii), the map $k[G] \to \mathrm{End}_{D_i}(V_i) \cong \prod_{i=1}^r \mathrm{M}_{n_i}(D_i^{op})$ is surjective.

Since the (ρ_i, V_i) exhaust all the simple left modules of $k[G]$, the kernel of this homomorphism equals $\mathrm{rad}\, k[G]$, which equals 0 as $k[G]$ is assumed to be semisimple. Thus, the map $k[G] \to \mathrm{End}_{D_i}(V_i) \cong \prod_{i=1}^r \mathrm{M}_{n_i}(D_i^{op})$ is an isomorphism, as required. □

Remark 19.24. (i) Thus, to every irreducible finite dimensional representation (ρ, V) of a group G over a field k, a division algebra D is attached:

$$D := \mathrm{End}_G(V) := \mathrm{End}_{k[G]}(V) = \{T : V \to V \ k \text{ linear } \mid T(g \cdot v) = g \cdot T(v) \,\forall\, g \in G\}.$$

This division algebra has the property that $\rho(k[G]) \subset \mathrm{End}_k(V)$ equals $\mathrm{End}_D(V)$, and $\rho : G \to GL_k(V)$ itself can be factored as a representation

$$\rho : G \to GL_D(V) \subset GL_k(V).$$

This division algebra, being canonically defined, is determined up to isomorphism.

(ii) In Burnside's theorem – either of (iv) or (v) of Corollary 19.17 – the assumption that k is algebraically closed is necessary. Indeed, if (ρ, V) is an irreducible representation of G over a field k, the image of $\rho : k[G] \to \mathrm{End}_k(V)$ is $\mathrm{End}_D(V)$, where D is the division algebra attached to ρ as above. When k is not algebraically closed, D can strictly contain k: see Example 19.25(i) below.

Example 19.25. (i) Here is an example of a representation $\rho : G \to GL_2(\mathbb{R})$, where $\rho(\mathbb{R}[G])$ is not surjective. Set $V = \mathbb{R}^2$. Let

$$G = SO_2(\mathbb{R}) = \left\{g \in GL_2(\mathbb{R}) \mid g \cdot {}^t g = 1, \det g = 1\right\} = \left\{\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \mid \theta \in \mathbb{R}\right\},$$

and let $\rho : G \hookrightarrow GL_2(\mathbb{R})$ be the inclusion. Then it is easy to see that

$$\rho(\mathbb{R}[G]) = \left\{r \cdot \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \mid r, \theta \in \mathbb{R}\right\} = \cong \left\{\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a, b \in \mathbb{R}\right\} \cong \mathbb{C},$$

using the map that sends $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ to $a + ib$.

Show as an easy exercise that the division algebra $D = \mathrm{End}_{\mathbb{R}[G]}(V)$ is isomorphic to $\mathbb{C}$ in this case.

Hint: $\rho(\mathbb{R}[G]) \subset \mathrm{End}_{\mathbb{R}}(V)$ is is isomorphic to $\mathbb{C}$, so as a $\rho(\mathbb{R}[G])$-module, $V = \mathbb{R}^2$ has dimension 1, so $\mathrm{End}_{\mathbb{R}[G]}(V)$ is isomorphic to $\mathbb{C}$.

(ii) Burnside's theorem implies that $Sp_{2n}(\mathbb{C})$ spans $M_{2n}(\mathbb{C})$: this is because $Sp_{2n}(\mathbb{C}) \hookrightarrow GL_{2n}(\mathbb{C}) \circlearrowright \mathbb{C}^{2n}$ is an irreducible representation, which in turn follows from the fact that $Sp_{2n}(\mathbb{C})$ acts transitively on $\mathbb{C}^{2n} \backslash \{0\}$ (any nonzero vector in $\mathbb{C}^{2n}$ can be extended into a symplectic basis). Slightly modifying this argument, one can see that $O_n(\mathbb{C}) \hookrightarrow GL_n(\mathbb{C})$ is an irreducible representation as well, so $O_n(\mathbb{C})$ spans $M_n(\mathbb{C})$.

(iii) In fact, $Sp(V,B)$ spans $\mathrm{End}_k(V)$, for any field k and any (finite dimensional) symplectic space (V,B) over k: to see this, note that the image of $k[Sp(V,B)] \to \mathrm{End}_k(V)$ is $\mathrm{End}_D(V)$ for some division algebra D, and it is not hard to see that the centralizer of $Sp(V,B)$ in $\mathrm{End}_k(V)$ is just the subring $k \subset \mathrm{End}_k(V)$ of scalar linear transformations. This would not be true with $\mathrm{SO}(V,q)$, as we saw with $\mathrm{SO}(2,\mathbb{R})$ above.

(iv) We have already found irreducible representations $\rho : G \to GL_{\mathbb{R}}(V)$, such that the associated division $\mathbb{R}$-algebra $\mathrm{End}_{\mathbb{R}[G]}(V)$ is $\mathbb{R}$ or $\mathbb{C}$. Thus, one might ask if we can find an irreducible representation $\rho : G \to GL_{\mathbb{R}}(V)$ for which the associated division algebra $\mathrm{End}_{\mathbb{R}[G]}(V)$ is the Hamilton quaternions $\mathbb{H}$. We think of $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}k$ as $\mathbb{C} \oplus \mathbb{C}j$, and consider $G := Q_8 := \{\pm 1, \pm i, \pm j, \pm k\}$, a subset of $\mathbb{H}$ closed under multiplication. We consider the representation

$$G \to GL_{\mathbb{R}}(\mathbb{H}) \cong GL_4(\mathbb{R}),$$

sending $g \in G$ to left multiplication by g on $\mathbb{H}$. Then $\rho(\mathbb{R}[G])$ is the $\mathbb{R}$-span, in $\mathrm{End}_{\mathbb{R}}(\mathbb{H})$, of left-multiplications by $\pm 1, \pm i, \pm j, \pm k$, namely, the collection of all left-multiplications by elements of $\mathbb{H}$. Thus, $\mathrm{End}_{\mathbb{R}[G]}(\mathbb{H})$ is the collection of all right-multiplications by elements of $\mathbb{H}$, namely, $\mathbb{H}^{op} \cong \mathbb{H}$ (show that $a + bi + cj + dk \mapsto a - bi - cj - dk$ defines an isomorphism $\mathbb{H}^{op} \to \mathbb{H}$). The irreducibility of this representation follows from the isomorphism $\mathrm{End}_{\mathbb{R}[G]}(\mathbb{H}) \cong \mathbb{H}$ and the semisimplicity of the representation; we will see in Lecture 20 that this semisimplicity follows in turn from the fact that G is finite and $\mathrm{char}\,\mathbb{R} = 0$.

As an aside, we now state von Neumann's double centralizer/bicommutant theorem, though this probably really belongs to the previous subsection:

Theorem 19.26 (von Neumann)**.** *Let $B(H)$ be the ring of bounded linear operators on a Hilbert space H, and $R \subset B(H)$ a subalgebra that is closed under $A \mapsto A^*$. Let $R' \subset B(H)$ be the centralizer of R, and $R'' \subset B(H)$ the centralizer of R', i.e., the double centralizer of R. Then R is dense in R'', in the sense that R'' is the closure of R under the weak operator topology on $B(H)$, and also under the strong operator topology on $B(H)$.*

19.6. **Rieffel's double centralizer theorem and simple rings.** The results of this subsection were only stated in the lecture.

Rieffel's theorem gives a completely different situation, in comparison with the Jacobson density theorem, under which $R \to R''$, as in the set up for the density/double centralizer theorems (see Subsection 19.2), is an isomorphism: it involves a simple ring rather than an arbitrary ring, and it involves a nonzero (possibly non-semisimple) left ideal rather than a semisimple module. Neither theorem implies the other, but if I understand Professor Nair's notes right, Section XVII.7 of Lang's book seems to give a common generalization.

After discussing the theorem, we will give conditions under which a simple ring is semisimple. Recall that a simple ring is one that is nonzero and has no proper nonzero two-sided ideals: it need not be semisimple (Professor Nair's notes give a simple ring that is not

semisimple, which is copied into HW 9). Recall also that this notation is different from that in Lang's book. The following proof is copied from Professor Nair's notes.

Theorem 19.27 (Rieffel's theorem)**.** *Let R be a simple ring and $I \subset R$ a nonzero left ideal. Form $R' = \mathrm{End}_R(I)$ and $R'' = \mathrm{End}_{R'}(I)$. Then $R \to R''$ is an isomorphism.*

Proof. The kernel of $\lambda : R \to R''$ is a proper two-sided ideal, and is hence zero. Therefore,it is enough to show that $R \to R''$ is surjective.

We claim that $\lambda(I) \subset R''$ is a left ideal. If this is granted, then we have, using that $IR = R$ (since $IR \subset R$ is a nonzero two-sided ideal):

$$R'' = R''\lambda(R) = R''\lambda(IR) = R''\lambda(I)\lambda(R) = \lambda(I)\lambda(R) \subset \lambda(R),$$

as desired.

To show that $\lambda(I) \subset R''$ is a left ideal, it is enough to show that for all $i \in I$ and $r'' \in R''$, we have $r''\lambda(i) = \lambda(r'' \cdot i)$, where $r'' \cdot i$ is made sense of by using that I is a left R''-module by the definition of $R'' \subset \mathrm{End}_{\mathbb{Z}}(I)$.

For all $j \in I$, viewing right multiplication by j as an endomorphism ρ_j of I that belongs to R', we have

$$r''\lambda(i)(j) = r''(ij) = r''(\rho_j(i)) = \rho_j(r''(i)) = r''(i)j = \lambda(r'' \cdot i)(j),$$

so that $r''\lambda(i) = \lambda(r'' \cdot i) \in \lambda(I)$, as desired. □

Theorem 19.28 (Wedderburn)**.** *If R is a simple ring, the following are equivalent:*

(i) R has a minimal nonzero left ideal.
(ii) R is left Artinian.
(iii) R is semisimple.
(iv) $R \cong M_n(D)$ for some division ring D and a positive integer n.

Proof. (iii) $\iff$ (iv) follows from the structure theorem for semisimple rings by Artin and Wedderburn. (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) is immediate/seen before. It is now enough to show that (i) implies (iv), which is what requires Rieffel's theorem.

Let $I \subset R$ be a minimal nonzero left ideal. By Rieffel's theorem we have $R \cong \mathrm{End}_{R'}(I) = \mathrm{End}_D(I)$, where $D = R' = \mathrm{End}_R(I)$ is a division algebra, because I being minimal nonzero is a simple R-module.

To finish the proof, it is enough to show that I is finite dimensional as a left D-vector space. But if not, the subset of $R = \mathrm{End}_D(I)$ consisting of finite rank operators would be a proper nonzero two-sided ideal, contradicting that R is simple. Therefore, $\dim_D I = n < \infty$ for some n, and $R \cong M_n(D)$. □

19.7. **Central simple algebras and Brauer groups – impressionistic introduction.** On this topic, mostly only definitions were given in the lecture. So the proofs in this section are optional reads.

Lemma 19.29. *Let A be a finite dimensional simple k-algebra. Then*

(i) $A \cong M_n(D)$ as a k-algebra, where D is a finite dimensional division k-algebra (i.e., $\dim_k D < \infty$; it is automatic that k sits inside $M_n(D)$ as scalar matrices).
(ii) The center of A is a finite field extension of k contained in D.

Proof. Since $\dim_k A < \infty$, A is Artinian in addition to being simple, so by Theorem 19.28, A is semisimple as well.

Hence by the theorem of Artin and Wedderburn, $A = M_n(D)$ for a division ring D. k lies in the center of $A = M_n(D)$, which identifies with the center $Z(D)$ of D (Exercise 19.1(iii)). Since D is a finite dimensional k-algebra, so is $Z(D) \subset D$. Since it is easy to see that $Z(D)\backslash\{0\}$ is closed under $x \mapsto x^{-1}$, we have that $Z(D)$ is a finite extension of k. □

Definition 19.30. Let k be a field. A central simple algebra over k is a finite dimensional k-algebra which is simple, and whose center is k (rather than a proper finite extension of k; this condition is the 'central' of a 'central simple algebra').

Remark 19.31. Thus, by Lemma 19.29, we can also define a central simple algebra over k as a k-algebra isomorphic to $M_n(D)$, where D is a central division algebra over k.

Lemma 19.32. *(i) Let A, A' be two simple k-algebras such that A is finite and central over k. Then $A \otimes_k A'$ is simple.*
(ii) If A, A' are central simple algebras over k, then $A \otimes_k A'$ is a central simple algebra over k.

Proof. Omitted: see https://stacks.math.columbia.edu/tag/074F for a proof of the first assertion, and https://stacks.math.columbia.edu/tag/074G for a proof of the second. The proofs do not need the introduction of new tools, but there is some argument that is to be made. □

Remark 19.33. Lemma 19.32 shows how important the condition of being 'central' is: for example, $\mathbb{C}$ is a finite simple algebra over $\mathbb{R}$, but not central, and $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \times \mathbb{C}$ is not simple.

We define a 'multiplication' on the set of isomorphism classes of central simple algebras over a field k, by defining the product of A and A' to be $A \otimes_k A'$, which is a central simple algebra over k by Lemma 19.32 above.

It is clear that this multiplication is associative, and has a multiplicative identity, namely the isomorphism class of the k-algebra k. Thus, the isomorphism classes of central simple algebras over k form a monoid.

However, this is too large and we would also like a group structure rather than a monoid structure. So we define:

Notation 19.34. Given central simple algebras A, A' over k, define $A \sim A'$ if $M_m(A) \cong M_n(A')$ for some positive integers m and n. Write $[A]$ for the equivalence class of A.

Exercise 19.35. (i) Show that $\sim$ is an equivalence relation, and each equivalence class contains exactly one division algebra.
Hint: We saw in Lecture 18 that D can be recovered from $R = M_n(D)$ as the opposite ring of $\mathrm{End}_R(M)$, where M is the unique simple R-module.
(ii) Show that $M_m(A) \otimes_k M_n(A') \cong M_{mn}(A \otimes_k A')$, and hence show that the multiplication defined above descends to one on the set of equivalence classes of central simple algebras over k.

Notation 19.36. For any field k, the Brauer group $Br(k)$ of k is the monoid of equivalence classes for $\sim$, under the operation $[A] \cdot [A'] = [A \otimes_k A']$ (see Exercise 19.35(ii) above). It follows from Lemma 19.37 below (as explained in Remark 19.38 below) that $Br(k)$ is a group, and not just a monoid.

Lemma 19.37. *If A is a central simple algebra over a field k of dimension n, we have $A \otimes_k A^{op} \cong \mathrm{End}_k(A) \cong M_n(k)$.*

Remark 19.38. It follows from Lemma 19.37 that in $Br(k)$, $[A]$ has an inverse given by $[A^{op}]$, so that $Br(k)$ is a group and not just a monoid.

Proof of Lemma 19.37. We have morphisms $A \to \mathrm{End}_k(A)$ and $A^{op} \to \mathrm{End}_k(A)$ defined by left and right multiplications (see Example 19.16(iii)). The images of these maps commute, so by the universal property of tensor product (see Exercise 19.10), we get a k-algebra homomorphism $A \otimes_k A^{op} \to \mathrm{End}_k(A)$.

We know that $A \otimes_k A^{op}$ is a central simple algebra, by Lemma 19.32 (this was the argument missing/postponed in Example 19.16(iii)). Therefore, $A \otimes_k A^{op} \hookrightarrow \mathrm{End}_k(A)$ is injective. Since both sides have dimension equal to $(\dim_k A)^2$, this map is also surjective. □

Theorem 19.39 (Skolem-Noether)**.** *Let A be a central simple algebra over k, let B be a simple k-algebra, and let $f, g : B \to A$ be two nonzero k-algebra homomorphisms. Then there exists an invertible $x \in A$ such that for all $b \in B$, we have $f(b) = xg(b)x^{-1}$.*

Before proving this result, let us observe a corollary which is not at all obvious, even when $A = M_n(k)$.

Corollary 19.40. *Any automorphism of a central simple algebra A is inner, i.e., given by $y \mapsto xyx^{-1}$, for some invertible $x \in A$.*

Proof. Take $B = A$ in Theorem 19.39. □

Proof of Theorem 19.39. Since A is simple and semisimple, it has a unique isomorphism class of simple left modules. Let M be a simple left A-module, so that $D = \mathrm{End}_A(M)$ is a division algebra over k. It is easy to see that $\dim_k M < \infty$, and then that $\dim_k D < \infty$.

Note that M is a left $A \otimes_k D$-module, and hence a left $B \otimes_k D$-module in two different ways: one via $f \otimes \mathrm{id}_D : B \otimes_k D \to A \otimes_k D$, and another via $g \otimes \mathrm{id}_D : B \otimes_k D \to A \otimes_k D$.

Since B is a simple nonzero algebra over k, we have by Lemma 19.32 that $B \otimes_k D$ is a simple k-algebra. Since B is simple and $f, g : B \to A$ are nonzero, we have $\dim_k B \otimes_k D < \infty$ as well, so that $B \otimes_k D$ is semisimple. Being simple and semisimple $B \otimes_k D$ has only one isomorphism class of simple modules. Thus, the two left $B \otimes_k D$-module structures on M are isomorphic: $B \otimes_k D$ is semisimple, and for either of the two left module structures, M has length equal to $\dim_k M / \dim_k N$, where N is any simple left $B \otimes_k D$-module.

Therefore, there exists an automorphism $\varphi : M \to M$ intertwining the two left $B \otimes_k D$-module structures on M, so that in particular we have:

$$\varphi(f(b) \cdot d \cdot m) = g(b) \cdot d \cdot \varphi(m),$$

for all $b \in B, d \in D$ and $m \in M$.

In particular, $\varphi : M \to M$ commutes with the action of D, so by the Jacobson density theorem and the fact that $\dim_k M < \infty$, φ is given by left multiplication by some $x \in A$. Thus, we have, for all $b \in B$ and $m \in M$:

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

Thus, left multiplication by $x \cdot f(b) \in A$ and $g(b) \cdot x \in A$ define the same endomorphism of M.

Since A is simple, $A \to \mathrm{End}_k(M)$ is injective, and therefore we get $x \cdot f(b) = g(b) \cdot x \in A$.

To conclude, it suffices to show that x is invertible in A. This is because $A \hookrightarrow \mathrm{End}_k(M)$ has image $\mathrm{End}_D(M)$ by the Jacobson density theorem, and $\varphi : M \to M$, which is induced by left-multiplication by x, is an isomorphism of D-vector spaces. □

There are many more lemmas related to central simple algebras of interest, which unfortunately we don't have time/space to discuss. We will conclude by giving some examples without proof. For this we will need the following fact too:

Remark 19.41. If K/k is a field extension, and A is a central simple algebra over k, one can show that $A \otimes_k K$ is a central simple algebra over K (see https://stacks.math.columbia.edu/tag/074H). It is now easy to see that $A \mapsto A \otimes_k K$ induces a group homomorphism $Br(k) \to Br(K)$.

Example 19.42.

(i) If k is algebraically closed, $Br(k)$ is trivial: this is because each equivalence class in $Br(k)$ contains exactly one division algebra, and over an algebraically closed field, there is no finite dimensional division algebra of dimension greater than 1 (Lemma 19.18).

(ii) If k is finite, $Br(k)$ is trivial: this follows from a theorem of Wedderburn, that any finite division algebra is commutative, and hence a field.

(iii) If $k = \mathbb{R}$, one can show that $Br(k) \cong \mathbb{Z}/2\mathbb{Z}$: the nontrivial division algebras in $Br(k) = Br(\mathbb{R})$ are $k = \mathbb{R}$ itself, and the Hamilton quaternions $\mathbb{H}$.

(iv) If $k = \mathbb{Q}_p$, with p a finite prime, one can show that $Br(k) \cong \mathbb{Q}/\mathbb{Z}$: this involves local class field theory.

(v) If $k = \mathbb{Q}$, then $Br(k) = Br(\mathbb{Q})$ is more complicated: Applying Remark 19.41 to $\mathbb{Q} \hookrightarrow \mathbb{R}$ and the inclusions $\mathbb{Q} \hookrightarrow \mathbb{Q}_p$, one can show that the sequence below is well-defined:

$$0 \to Br(\mathbb{Q}) \to \left(\bigoplus_p Br(\mathbb{Q}_p)\right) \oplus Br(\mathbb{R}) \cong \left(\bigoplus_p \mathbb{Q}/\mathbb{Z}\right) \oplus ((1/2)\mathbb{Z})/\mathbb{Z} \xrightarrow{\Sigma} \mathbb{Q}/\mathbb{Z} \to 0$$

(the obvious map is from $Br(\mathbb{Q})$ to $\prod_p Br(\mathbb{Q}_p) \times Br(\mathbb{R})$, and the well-definedness involves checking that this actually lands inside the direct sum). Using class field theory, one can see that the above sequence is exact, so that $Br(\mathbb{Q})$ is a kernel of a surjective map $(\bigoplus \mathbb{Q}/\mathbb{Z}) \oplus ((1/2)\mathbb{Z}/\mathbb{Z}) \to \mathbb{Q}/\mathbb{Z}$, where $\bigoplus \mathbb{Q}/\mathbb{Z}$ is an infinite sum.

Again, as with quadratic forms, we see that $\mathbb{Q}$ is much more complicated in many ways than $\mathbb{R}$ or the $\mathbb{Q}_p$, and that objects associated to $\mathbb{Q}$ are often number theoretic in nature.

There is much more very basic stuff to discuss concerning Brauer groups, for which we don't have the space or time. The 10 page pdf at https://stacks.math.columbia.edu/download/brauer.pdf is a good summary, with proofs, of various basic properties that we haven't discussed.

Or you can refer to Professor Nair's notes, which also have lots of lemmas and theorems about Brauer groups that we haven't discussed, and also some proofs that we have skipped: it all takes only four pages in his notes.

20. Lecture 20 – Representation theory of finite groups – I

In this lecture, unless otherwise mentioned, R denotes a commutative ring. Recall that $R[G]$ denotes the group ring of G, and that the category $Rep_R(G)$ of R-modules with a G-action identifies with the category $R[G]$-Mod of left $R[G]$-modules. A representation of G over R will often simply be referred to as a G-module, and we may write Hom_G for $\mathrm{Hom}_{R[G]}$.

20.1. Semisimplicity of $Rep_k(G)$.

Notation 20.1. (i) Henceforth, we will sometimes also work with a different but equivalent description of $R[G]$: identifying $\sum_{g\in G} a_g g$ with the map $G \to R$ given by $g \mapsto a_g$, we may think of $R[G]$ as the free R-module of finitely supported functions $G \to R$. Under this description, the ring multiplication we have already defined on $R[G]$ (using the 'group algebra' description of $R[G]$) translates to convolution:

$$(f_1 * f_2)(g) = \sum_{\substack{g_1,g_2\in G\\ g_1g_2=g}} f_1(g_1)f_2(g_2).$$

(ii) The regular representation associated to G over R is the representation of $G \times G$ on $R[G]$ defined by $(g_1, g_2)(\sum a_g g) = \sum_g a_g(g_1 g g_2^{-1})$, or equivalently by $((g_1, g_2) \cdot f)(g) = f(g_1^{-1} g g_2)$. The left regular representation of G is the restriction of the regular representation along $G \cong G \times \{1\} \hookrightarrow G \times G$ (thus, defined by $g_1 \cdot (\sum_g a_g g) = \sum_g a_g(g_1 g)$ or $g_1 \cdot f = f(g_1^{-1} \cdot -)$), and the right regular representation the restriction along $G \cong \{1\} \times G \hookrightarrow G \times G$ (thus, defined by $(\sum_g a_g g)g_2 = \sum_g a_g(g g_2)$ or $g_2 \cdot f = f(- \cdot g_2)$).

Remark 20.2. Pretend for a moment that $R = \mathbb{R}$ or $\mathbb{C}$, so that we can talk of (signed) real or complex measures on the discrete group G. The group algebra should 'really' be viewed as the space of finitely supported such measures on G, made into a ring via convolution: the convolution of two measures μ_1 and μ_2 on (the discrete space) G is the push-forward, along $G \times G \to G$, of the product measure $\mu_1 \times \mu_2$ on $G \times G$. This is more natural because while functions naturally pull back, measures naturally push forward. When we identify $R[G]$ with the space of finitely supported functions $G \to R$, as in Notation 20.1(i), we are really taking the Radon-Nikodym derivative with respect to the counting measure.

A similar comment applies even to more general R: you can imitate measures by thinking of $\sum_{g\in G} a_g g$ as the map

$$\{\text{Functions } G \to k\} \to k,$$

sending f to $\sum_{g\in G} a_g f(g)$ – this is viewed as integrating f against $\sum_g a_g \delta_g$, where δ_g is the Dirac delta measure at g.

Definition 20.3. For any group G, the map $\varepsilon : R[G] \to R$ given by

$$\sum_{g\in G} a_g g \mapsto \sum_{g\in G} a_g,$$

is easily verified to be a homomorphism, and is called the augmentation map. The two-sided ideal $\ker \varepsilon \subset R[G]$ is called the augmentation ideal.

When we talk of the augmentation map, the $\sum_{g\in G} a_g g$ description of $R[G]$ will be more convenient. Note that the augmentation map intertwines the regular representation of $G \times G$ on $R[G]$ with the trivial representation of $G \times G$ on R: $\varepsilon((g_1, g_2) \cdot f) = \varepsilon(f)$ for all $g_1, g_2 \in G$. Similarly, it intertwines the left and the right regular representations of G with the trivial representation of G on R.

Proposition 20.4. *Given a nonzero commutative ring R and a group G, the following are equivalent:*

(i) $R[G]$ is semisimple.
(ii) R is semisimple, and the ideal $\ker \varepsilon \subset R[G]$ has a complementary left ideal, i.e., there exists a left ideal $J \subset R[G]$ such that $R = (\ker \varepsilon) \oplus J$.
(iii) R is semisimple, G is finite, and $\#G \in R^\times$ (i.e., the image of $\#G \in \mathbb{Z}$ under $\mathbb{Z} \to R$ belongs to $R^\times$).

Remark 20.5. Note that the condition (iii) of the proposition translates to the following: R is a finite product $k_1 \times \cdots \times k_r$ of fields, and $(\operatorname{char} k_i, \#G) = 1$ for all $1 \leqslant i \leqslant r$.

Corollary 20.6 (Maschke)**.** *If G is a finite group and k is a field with $(\operatorname{char} k, \#G) = 1$, then $k[G]$ is semisimple.*

Proof of Proposition 20.4. It is easy to see that if $S_1 \to S_2$ is a surjective ring homomorphism and S_1 is semisimple, then S_2 is semisimple as well (if a left S_2-module is completely reducible as a left S_1-module, then it is completely reducible as a left S_2-module as well). Applying this to $\varepsilon : R[G] \to R$, we conclude that if $R[G]$ is semisimple then so is R. Now it is immediate that (i) implies (ii): the existence of J is a very special case of semisimplicity of $R[G]$.

Now we prove (ii) $\Rightarrow$ (iii). Giving $R[G]$ the left regular action of G, ε restricts to a left $R[G]$-module isomorphism $J \to R$. It follows that G fixes $J \subset R[G]$ pointwise. But the only elements $\sum_{g\in G} a_g g \in R[G]$ that can be fixed under left multiplication by G are those for which $a_g = a_h$ for all g, h. Since $\sum_{g\in G} a_g g$ is a finite sum, and since J is nonzero (as ε is surjective and $R \neq 0$), it follows that G is finite. This also gives that $J \subset R \cdot (\sum_{g\in G} g)$, so $R(\sum_{g\in G} g) + (\ker \varepsilon) = R[G]$. Applying ε, we get that $(\#G) \cdot R = R$, so that $\#G \in R^\times$.

Now let us prove (iii) $\Rightarrow$ (i). It is enough to show that any $R[G]$-submodule $\iota : V' \hookrightarrow V$ of a left $R[G]$-module V has a complementary left $R[G]$-module, or equivalently a G-equivariant section $p : V \to V'$. Since R is semisimple, ι has an R-linear section $b : V \to V'$, and we wish to replace it with an $R[G]$-linear section $p : V \to V'$, one that is R-linear and satisfies $g \cdot p \cdot g^{-1} = p$ for all $g \in G$. We may then take

$$p(v) = \frac{1}{\#G} \left(\sum_{g\in G} (g \cdot b \cdot g^{-1})(v) \right).$$

It is immediate that p is an $R[G]$-linear section to ι, as desired. $\square$

Before making several remarks on the proof of Proposition 20.4, let us define some constructs:

Notation 20.7. (i) Often, for $V \in \mathrm{Ob}\, Rep_R(G) = R[G]\text{-}Mod$, we will denote by $V^G \subset V$ the R-submodule of G-fixed elements of V.

(ii) If $V_1, V_2 \in \mathrm{Ob}\, Rep_R(G) = R[G]\text{-}Mod$, then $\mathrm{Hom}_R(V_1, V_2)$ is a priori an R-module, but can also be viewed as a left $R[G]$-module, where for $g \in G$ and $\varphi \in \mathrm{Hom}_R(V_1, V_2)$, $g \cdot \varphi \in \mathrm{Hom}_R(V_1, V_2)$ is defined by:

$$(g \cdot \varphi)(v_1) = g \cdot \varphi(g^{-1} \cdot v_1).$$

Thus, $\mathrm{Hom}_R(V_1, V_2)^G = \mathrm{Hom}_{R[G]}(V_1, V_2) \subset \mathrm{Hom}_R(V_1, V_2)$. Note the relevance of this construction to the proof of (iii) $\Rightarrow$ (i) of Proposition 20.4: it helped reduce the construction of a G-invariant homomorphism to that of a G-invariant element in a module.

The observartion that $\mathrm{Hom}_R(V_1, V_2)$ is a G-module can be formalized by saying that $R[G]\text{-}Mod$ has an 'internal hom': one can associate to objects X, Y in this abelian category another object of the same category (rather than a set or an abelian group) called $\underline{Hom}(X, Y)$, which behaves in some ways like Hom between X and Y. Please note that $\mathrm{Hom}_{R[G]}(V_1, V_2)$ is still only an abelian group, and is much smaller than the left $R[G]$-module $\mathrm{Hom}_R(V_1, V_2)$.

(iii) If $V_2 = R$ is the trivial representation of G, then we denote $\mathrm{Hom}_R(V_1, V_2)$ – viewed as a representation of G – by $V_1^\vee$, and call it the contragredient representation of V_1.

(iv) As with Hom, so with $\otimes$: if $V_1, V_2 \in \mathrm{Ob}\, Rep_R(G) = R[G]\text{-}Mod$, the R-module $V_1 \otimes_R V_2$ can be upgraded to a left $R[G]$-module by defining $g \cdot (v_1, v_2) = (g \cdot v_1, g \cdot v_2)$.

This is essentially what makes $Rep_R(G) = R[G]\text{-}Mod$ what is called a tensor category or a monoidal category, something that is related to realizing the 'internal hom's mentioned above (by imposing an 'internal' Hom-tensor adjointness).

Remark 20.8. (i) Here is an easy exercise related to the theme of semisimplicity: show that a ring S is semisimple if and only if every left S-module is projective (similarly, S is semisimple if and only if every left S-module is injective).

(ii) If R equals $\mathbb{R}$ or $\mathbb{C}$, the proposition tells us that for any finite group G, $R[G]$ is semisimple. In this case, an alternate justification for this fact can be given, as follows. Choose any inner product (real or complex, as applicable) $(\cdot, \cdot)$ on V, and replace it with the inner product $\langle \cdot, \cdot \rangle$ given by

$$\langle \cdot, \cdot \rangle = (\#G)^{-1} \sum_{g \in G} g \cdot (\cdot, \cdot),$$

where of course $g \cdot (\cdot, \cdot) = (g \cdot -, g \cdot -)$, to get a G-invariant inner product on V. Since $V' \subset V$ is nondegenerate (for an inner product, by positive definiteness all subspaces are nondegenerate), the orthogonal $(V')^\perp \subset V$ is really an orthogonal

complement, i.e., satisfies $V' \oplus (V')^{\perp} = V$. It is easy to see from the G-invariance of the inner product that $(V')^{\perp}$ is stable under G, and thus gives an $R[G]$-module complement to V'.

This proof is particular to $\mathbb{R}$ or $\mathbb{C}$, but shows, for instance, that any finite subgroup of $GL_n(\mathbb{R})$ can be conjugated into $O_n \subset GL_n(\mathbb{R})$, and that any finite subgroup of $GL_n(\mathbb{C})$ can be conjugated into the unitary group U_n. Further, this idea seems to be of importance to functional analysis, where 'semisimplicity' seems to often be ensured using self-adjointness assumptions.

(iii) The idea of Maschke's proof (namely the proof of (iii) $\Rightarrow$ (i) of the proposition) also adapts to representations of compact topological groups G, where rather than arbitrary representations one considers appropriately continuous representations. In this situation, the averaging operation $(\#G)^{-1}\sum_{g\in G}$ can be replaced by the operation $\int_G(\cdot)\,dg$ of integrating against the normalized Haar measure on G.

This also applies to the argument of the above point, where one produces G-invariant inner products, and shows that any compact subgroup of $GL_n(\mathbb{R})$ can be conjugated to one contained in O_n, and any compact subgroup of $GL_n(\mathbb{C})$ can be conjugated to one contained in U_n. Thus, O_n is the unique conjugacy class of maximal compact subgroups of $GL_n(\mathbb{R})$, while U_n is the unique conjugacy class of maximal compact subgroups of $GL_n(\mathbb{C})$. In the theory of real reductive groups, one generalizes this to other 'reductive' groups such as symplectic and special orthogonal groups over $\mathbb{R}$ and $\mathbb{C}$.

(iv) We now outline another way of describing Maschke's proof, assuming for simplicity that k is a field: $(\operatorname{char} k, \#G) = 1$, and we want to show that $k[G]$ is semisimple. I learnt about this sort of a proof from Professor Nair, and while this articulation is more tedious, here are somethings I like about it:

- It seems to give some insight into why (ii) of Proposition 20.4, though a special case of (i), implies the whole of (i), and as to why the technique of averaging in (iii), which one might a priori expect to only give sections to $V^G \hookrightarrow V$, also gives sections to more general $V' \to V$. While it seems trivial to read and formally verify Maschke's proof, it seems to have some mystery as to why, philosophically, a special case implies the general case.
- It seems to much better parallel the proof of semisimplicity of representations of semisimple Lie algebras in characteristic zero – if you don't appreciate this now, you can skip this, and come back to it say in a later semester when you read semisimple Lie algebras.

Here are the steps, written partly as exercises, where we now assume that $(\operatorname{char} k, \#G) = 1$:

(a) It is also enough to a show that any epimorphism $V \to V''$ in $Rep_k(G)$ has a section (to find a complement to the image of $\iota : V' \hookrightarrow V$, apply this to the epimorphism $V \to V/V'$).

(b) Consider the obvious surjection $\operatorname{Hom}_k(V'', V) \to \operatorname{Hom}_k(V'', V'')$, look at the preimage of $k \cdot \mathrm{id}_{V''} \subset \operatorname{Hom}_{k[G]}(V'', V'')$, which is a G-stable subspace of

$\mathrm{Hom}_k(V'', V)$, and use this to reduce the semisimplicity of $k[G]$ to the following assertion: any exact sequence

$$0 \to W \to V \to k \to 0$$

in $k[G]$-Mod, where k is viewed as the trivial representation of G, splits.

(c) Reduce further, using induction, to the case where W is an irreducible representation of G.
Hint: If $W_0 \subsetneq W$ is a proper nonzero G-submodule, we get an exact sequence $0 \to W/W_0 \to V/W_0 \to k \to 0$ of G-modules, which splits by induction, giving an exact sequence $0 \to W_0 \to V' \to k$ for some G-stable subspace V' of V.

(d) There are now two cases: one where W is the trivial representation of G, and one where W is nontrivial.
- First suppose W is nontrivial. Then $c := (\#G)^{-1} \cdot (\sum_{g \in G} g) \in R[G]$ maps W to $W^G = 0$, while it acts as the identity on $V/W = k$. Thus, $c-1$, which is a central element of $R[G]$, annihilates k but not W, and it satisfies $(c-1)V \subset W$ and $c-1|_W = -1$. Then $\ker(c-1)$ is the desired complement.
- If W is trivial, then for all $g \in G$, $(g-1)$ takes W to 0 and V to W, so $(g-1)^2$ annihilates V. But $g^{\#G} - 1$ also annihilates V. Since $x^{\#G-1}$ and $(x-1)^2$ are relatively prime in $k[x]$, by the assumption that $(\mathrm{char}\, k, \#G) = 1$, it follows that $g-1$ annihilates V for all $g \in G$, i.e., V is the trivial representation. Then, any k-linear splitting will do.

In the Lie algebra case, the role of $c-1$ seems to be played by the Casimir element: it is to bring this out that the above was made harder than it should be, without using that c is an idempotent. Note that c is a central element of $R[G]$, and the above proof is a "central character argument", along the lines of "eigenspaces corresponding to distinct eigenvalues are linearly independent". The argument in the case where W is nontrivial can be articulated in one sentence if one understands Ext well enough, which formalizes the idea of the preceding sentence. See Professor Nair's comments in the proof he gives of the Lie algebra case.

20.2. **The group algebra and matrix algebras of irreducible representations.**

Notation 20.9. Whenever $\rho : G \to GL_k(V)$ is a representation of a group G, with V finite dimensional over a field k, we will let ρ also stand for the k-algebra homomorphism $\rho : k[G] \to \mathrm{End}_k(V)$ obtained by k-linearly extending the group homomorphism $\rho : G \to GL_k(V)$.

Theorem 20.10. *Let k be a field and G a finite group. Then G has only finitely many irreducible representations up to isomorphism, say $(\rho_1, V_1), \ldots, (\rho_r, V_r)$. Moreover:*

(i) The map

$$\rho = \prod_{i=1}^{r} \rho_i : k[G] \to \prod_{i=1}^{r} \mathrm{End}_k(V_i)$$

induces an isomorphism of k-algebras

$$k[G]/(\operatorname{rad} k[G]) \to \prod_{i=1}^{r} \operatorname{End}_{D_i}(V_i),$$

where for $1 \leqslant i \leqslant r$, $D_i = \operatorname{End}_G(V_i) := \operatorname{End}_{k[G]}(V_i)$ (it is a division algebra over k).

(ii) There is also an isomorphism

$$k[G]/(\operatorname{rad} k[G]) \cong \prod_{i=1}^{r} \operatorname{End}_{D_i}(W_i),$$

where $W_i = \operatorname{Hom}_{k[G]}(V_i, k[G])$ is the multiplicity space of V_i in $k[G]$, on which $D_i = \operatorname{End}_{k[G]}(V_i)$ operates on the right via its action on V_i (so W_i is a right D_i-vector space).

(iii) $\sum_{i=1}^{r} (\dim_{D_i} V_i)^2 \dim_k D_i = \sum_{i=1}^{r} (\dim_{D_i} W_i)^2 \dim_k D_i = \dim(k[G]/\operatorname{rad} k[G]) \leqslant \#G$, with equality if and only if $k[G]$ is semisimple, i.e., if and only if $(\#G, \operatorname{char} k) = 1$.

Proof. Recall that for any left Artinian ring R, $\operatorname{rad}(R/\operatorname{rad}(R)) = 0$, and that hence $R/\operatorname{rad}(R)$ is a semisimple ring with the same collection of simple left modules as R. Thus, simple left modules for $k[G]/(\operatorname{rad} k[G])$ identify with those for $k[G]$. The finiteness of the number of isomorphism classes of these modules was (an easy) part of the Artin-Wedderburn theorem (Lecture 18).

Now (i) is one of the consequences of Jacobson density theorem, that we studied in Lecture 19 – see Corollary 19.23 from Lecture 19 (it might also follow from the exercise in Lecture 18 that attempted an alternate take on the Artin-Wedderburn theorem).

(ii) follows from the isomorphism given by the Artin-Wedderburn structure theorem, provided we replace $k[G]/\operatorname{rad}(k[G])$ with $(k[G]/\operatorname{rad}(k[G]))^{op}$. Thus, to prove (ii), it is enough to show that $k[G]/\operatorname{rad}(k[G])$ is isomorphic to its own opposite. Since the identification $R \to R^{op}$ has been seen to preserve the Jacobson radical, it is enough to show that $k[G]$ is isomorphic to $k[G]^{op}$. Indeed, $g \mapsto g^{-1}$ defines an isomorphism $k[G] \to k[G]^{op}$.

(iii) follows from (i). □

Corollary 20.11. *Let $k = \bar{k}$ be an algebraically closed field and G a finite group. Let $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ be the finitely many irreducible representations of G up to isomorphism. Then $\prod_{i=1}^{r} \rho_i : k[G] \to \prod_{i=1}^{r} \operatorname{End}_k(V_i)$ quotients to an isomorphism*

$$k[G]/(\operatorname{rad} k[G]) \cong \prod_{i=1}^{r} \operatorname{End}_k(V_i),$$

so that

$$\sum_{i=1}^{r} (\dim_k V_i)^2 = \dim_k(k[G]/(\operatorname{rad} k[G])) \leqslant \#G,$$

with equality if and only if $k[G]$ is semisimple, i.e., if and only if $(\#G, \operatorname{char} k) = 1$.

Proof. Combine Theorem 20.10 with the fact that finite dimensional division k-algebras are all equal to k when k is algebraically closed – see Lemma 19.18 from Lecture 19. □

Remark 20.12. Note that a one-dimensional representation of G over a field k is a homomorphism $\rho : G \to GL_k(V) = GL_1(k) = k^\times$, where V is a one-dimensional vector space over k: here the isomorphism $GL_k(V) \to GL_1(k)$ is defined using the choice of a basis for V, but is easily seen to be independent of the choice of the basis since $\dim_k V = 1$.

20.3. **The abelian case.**

Corollary 20.13. *Let k be a field and G a finite abelian group. Assume notation from Theorem 20.10.*

(i) For each irreducible representation V of G over k, $\mathrm{End}_G(V)$ is a finite field extension of k. In other words, each division k-algebra D_i as in Theorem 20.10 is a finite field extension K_i/k, and each V_i is a one-dimensional K_i-vector space. Note that while we know that $k[G]/\operatorname{rad}(k[G])$ being commutative and semisimple is a finite product of fields, this gives an explicit such realization:

$$k[G]/\operatorname{rad}(k[G]) = \prod_{i=1}^{r} K_i.$$

(ii) Suppose further that $k = \bar{k}$ (and G is still abelian). Then each irreducible representation (ρ, V) of G over k is one-dimensional, and can hence be thought of as a homomorphism $\rho : G \to k^\times$ (see Remark 20.12). Hence using the notation of Theorem 20.10 we get an isomorphism

$$k[G]/\operatorname{rad}(k[G]) \cong \prod_{\chi \in \mathrm{Hom}(G, k^\times)} k, \tag{76}$$

induced by sending $g \in G \subset k[G]$ to $(\chi(g))_\chi$ (see Exercise 20.14(ii) below to spell this map out better).

Proof. Since G is abelian, it is easy to see that $k[G]$ is a commutative ring. Hence each $\mathrm{End}_{D_i}(V_i) \cong M_{n_i}(D_i^{op})$ is abelian, so that $D_i = K_i$ is a field extension of k and $n_i = \dim_{D_i} V_i = \dim_{K_i} V_i = 1$.

From this (i) follows, and (ii) is an immediate consequence. □

Exercise 20.14. (i) It follows from Corollary 20.13(ii) that if G is finite abelian and $k = \bar{k}$ is an algebraically closed field with $(\#G, \operatorname{char} k) = 1$, then any irreducible representation of G is one-dimensional. Show this directly using just Schur's lemma.

(ii) For any finite group G and field k, show that restriction along $G \hookrightarrow k[G]$ gives an isomorphism

$$\mathrm{Hom}_{k\text{-}Alg}(k[G], k) \to \mathrm{Hom}(G, k^\times).$$

(iii) (Important for understanding) Make sure you understand the following: the following is how you representation-theoretically interpret the simultaneous diagonalization of commuting diagonalizable matrices, in the case where they generate a finite group. Let $k = \bar{k}$ be algebraically closed, and let G be a finite abelian group such that $(\#G, \operatorname{char} k) = 1$.

(a) If (G, V) is any (possibly infinite dimensional) representation of G, since the irreducible representations of G are one-dimensional and hence (their isomorphism classes are) indexed by $\operatorname{Hom}(G, k^\times)$, the decomposition of the (necessarily semisimple) representation V of G into isotypic components takes the form:

$$V = \bigoplus_{\chi \in \operatorname{Hom}(G,k^\times)} V_\chi, \tag{77}$$

where $V_\chi \subset V$ is the subspace on which G acts by χ^{-1} (make sure you know what 'acts by χ^{-1}' means).

(b) Recall that a module V over the product ring $k[G] = \prod_{\chi \in \operatorname{Hom}(G,k^\times)} k$ is a direct sum

$$V = \bigoplus_{\chi \in \operatorname{Hom}(G,k^\times)} V_\chi, \tag{78}$$

where each V_{χ_0} is a module over $\prod_\chi k$ on which $\prod_\chi k$ acts through the projection $\prod_\chi k \to k$ onto the χ_0-th component. Then (78) is the same as (77) (that the latter had χ^{-1} rather than χ is something we will revisit in Lectures 21 and 22).

(c) What is the decomposition (77) or equivalently (78) when $k = k[G]$ (for simplicity, I will exchange χ with χ^{-1})? Show that the $\chi \in \operatorname{Hom}(G, k^\times) \subset k[G]$ give direct sum decomposition:

$$k[G] = \bigoplus_{\chi \in \operatorname{Hom}(G,k^\times)} k \cdot \chi. \tag{79}$$

Thus, $\operatorname{Hom}(G, k^\times) \subset k[G]$ is a k-vector space basis, different from $G \subset k[G]$: it is basis that diagonalizes the action of G on $k[G]$. Moreover, (79) is the isotypic decomposition of $k[G]$ for the left regular action of G (resp., the right-regular action of G; resp., the regular action of $G \times G$), with the k-span $k \cdot \chi \subset k[G]$ of $\chi \in \operatorname{Hom}(G, k^\times) \subset k[G]$ the isotypic component corresponding to the representation χ^{-1} of G (resp., the representation χ of G; resp., the representation $\chi^{-1} \otimes \chi$ of $G \times G$ given by $(g_1, g_2) \mapsto \chi(g_1^{-1} g_2)$).

In other words, an irreducible representation of a not necessarily abelian group G is the generalization, from the finite abelian algebraically closed case, of 'simultaneous diagonalization': an irreducible representation generalizes 'simultaneous eigenvalues', and hence sometimes the word 'spectrum' gets used to describe the set of isomorphism classes of irreducible representations.

Remark 20.15. Here is some comment on how we will continue on this theme in Lecture 21. Assume that $(\operatorname{char} k, G) = 1$. The $\chi \in \operatorname{Hom}(G, k^\times)$ when G is abelian should also remind you of the functions $z \mapsto z^n$ on S^1 or $x \mapsto e^{ixy}$ on $\mathbb{R}$ that you see in Fourier expansion: more on that in Lecture 21. In fact, the 'action map' $G \to \prod_{i=1}^r \operatorname{End}_{D_i}(V_i)$ can be thought of as a Fourier transform (and note that it takes convolution to multiplication, like a classical Fourier transform does), and the above exercise raises the question of how to 'Fourier invert' the isomorphism $k[G] \to \prod_{i=1}^r \operatorname{End}_{D_i}(V_i)$ from Theorem 20.10. At least when k is algebraically closed, we will hopefully discuss this in Lectures 21 and 22.

20.4. **Irreducible representations and conjugacy classes.**

Lemma 20.16. *Let k be a field. The center $Z(k[G])$ of $k[G]$ equals*

$$\Big\{\sum_g a_g g \mid a_g = a_{hgh^{-1}} \ \forall\, h, g \in G\Big\},$$

so that $\dim_k Z(k[G])$ is the number of conjugacy classes of G.

Proof. Since $k[G]$ is generated as a ring by k and $G \subset k[G]^\times$, the center of $k[G]$ is precisely the subspace fixed by the $\operatorname{Int} h$, as h varies over G. □

Proposition 20.17. *Assume that k is algebraically closed and that $(\operatorname{char} k, \#G) = 1$. Then the number of irreducible representations of G up to isomorphism is the number of conjugacy classes in G.*

Proof. Since each $\operatorname{End}_k(V)$ has center k, this follows from Lemma 20.16 and the isomorphism

$$k[G] \cong \prod_{i=1}^r \operatorname{End}_k(V_i)$$

– we have k instead of D_i since k is algebraically closed. □

Remark 20.18.
(i) The proof shows that if k is algebraically closed of characteristic p, and $(\operatorname{char} k, \#G) > 1$, then the number of irreducible representations is at most the number of conjugacy classes of G. A theorem of R. Brauer says that the number of irreducible representations of G up to isomorphism is the number of p-regular conjugacy classes of G, i.e., the number of conjugacy classes of G consisting of elements whose order is relatively prime to p.
(ii) On the other hand, in good characteristic, i.e., when $(\operatorname{char} k, \#G) = 1$, but where we do not assume that k is algebraically closed, the above proof shows that the number of irreducible representations of G is at most the number of conjugacy classes of G.

Exercise 20.19. Prove the aforementioned result of Brauer (mentioned in Remark 20.18(i)) in the case where G is abelian.

20.5. **Some examples.** First we show that irreducible representations of p-groups are trivial in characteristic p.

Proposition 20.20. *Suppose k is a field of characteristic $p > 0$, and that G is a p-group. Then every irreducible representation of G on a k-vector space is trivial.*

Proof. In the case where G is abelian and k is algebraically closed, we know that each irreducible representation of G is given by a character $\chi : G \to k^\times$, which is trivial as k has only one p-th root of unity. When k is algebraically closed, the general case is easy to deduce inductively from this, using that a p-group being nilpotent has a nontrivial center.

We now sketch an argument that does not need k to be algebraically closed, and directly takes care of a general (not necessarily abelian) p-group G. It is enough to show that each irreducible representation (ρ, V) of G contains a nonzero G-fixed vector. This immediately lets us reduce to the case where $k = \mathbb{F}_p$ (k contains a copy of $\mathbb{F}_p$, so consider the span over $\mathbb{F}_p[G] \subset k[G]$ of any nonzero vector in the representation). The G-orbits of non-fixed vectors all have cardinalities that are multiples of p. The cardinality of the vector space is also a multiple of p. So the set of G-fixed vectors should also have cardinality that is also a multiple of p, and cannot consist only of 0. □

Example 20.21. Let us study the irreducible representations of $G = S_3$ over any field k, which is not necessarily algebraically closed. Consider the representations $(\rho_1, V_1), (\rho_2, V_2)$ and (ρ_3, V_3) of G, where:

- (ρ_1, V_1) is the trivial representation of G,
- (ρ_2, V_2) is the 'sign character' $\operatorname{sgn} : S_3 \to \{\pm 1\}$ that sends a permutation to its sign (i.e., the determinant of the permutation matrix that represents this permutation and has only 0's and 1's as entries), and
- $V_3 = \{(a_1, a_2, a_3) \in k^3 \mid a_1 + a_2 + a_3 = 0\}$ and the action of S_3, via ρ_3, on $V_3 \subset k^3$ is by the obvious permutation of the coordinates.

Thus, $\dim_k V_1 = \dim_k V_2 = 1$ and $\dim_k V_3 = 2$. Note the following:

- ρ_1, ρ_2 and ρ_3 are distinct if $\operatorname{char} k \neq 2$, while $\rho_1 \cong \rho_2$ if $\operatorname{char} k = 2$; and
- ρ_1 and ρ_2 are clearly irreducible, while it is easy (exercise) to see that ρ_3 is irreducible if and only if $\operatorname{char} k \neq 3$.

From this, we conclude the following:

(i) Suppose $(\operatorname{char} k, \#S_3) = 1$, i.e., $\operatorname{char} k \neq 2, 3$. Then ρ_1, ρ_2 and ρ_3 are irreducible and pairwise distinct. Since S_3 has exactly three conjugacy classes, it follows from Remark 20.18(ii) (or Proposition 20.17 if k is algebraically closed) that $(\rho_1, V_1), (\rho_2, V_2)$ and (ρ_3, V_3) are exactly the irreducible representations of G up to isomorphism.

(ii) Suppose $\operatorname{char} k = 2$. Then (ρ_1, V_1) and (ρ_3, V_3) are distinct irreducible representations of G (though $\rho_1 \cong \rho_2$), so we have a surjection

$$\rho_1 \times \rho_3 : k[G]/(\operatorname{rad} k[G]) \twoheadrightarrow \operatorname{End}_k(V_1) \times \operatorname{End}_k(V_3).$$

Since the right-hand side has dimension 5 and the left-hand side has dimension strictly less than 6, we conclude that the above map is an isomorphism, so (ρ_1, V_1) and (ρ_3, V_3) are exactly the irreducible representations of G over k up to isomorphism.

(iii) Suppose $\operatorname{char} k = 3$. In this case, the normal subgroup $A_3 \subset S_3$ of order 3 fixes a nonzero vector in every irreducible representation of G by Proposition 20.20, and hence by normality acts trivially on every irreducible representation of G (please make sure you understand this deduction). Thus, the irreducible representations of S_3 over k identify with those of $S_3/A_3 \cong \mathbb{Z}/2\mathbb{Z}$, and hence are just the representations (ρ_1, V_1) and (ρ_2, V_2) (which are distinct since $\operatorname{char} k \neq 2$).

Exercise 20.22. Prove that the irreducible representations of $\mathbb{Z}/n\mathbb{Z}$ over an arbitrary field k are obtained as follows.

(i) If m is the largest factor of n such that the $\operatorname{char} k \nmid m$, then show that pull-back under $\mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ defines a bijection from the set of isomorphism classes of irreducible representations of $\mathbb{Z}/m\mathbb{Z}$ over k to the set of isomorphism classes of irreducible representations of $\mathbb{Z}/n\mathbb{Z}$ over k.

(ii) Thus, assume that $(n, \operatorname{char} k) = 1$. Let $f_1, \ldots, f_r$ be the irreducible factors of $x^n - 1$. For $1 \leqslant i \leqslant r$, let $K_i = k[x]/(f_i)$, and let $\alpha_i \in K_i$ be the image of x; it is an n-th root of 1 in the field K_i. Define a $[K_i : k_i]$-dimensional representation (ρ_i, V_i) of $G = \mathbb{Z}/n\mathbb{Z}$ as follows: $V_i = K_i$, viewed as a vector space over k, and $\rho_i : G \to GL_k(V_i) = GL_k(K_i)$ is the unique map that sends the generator $\bar{1} \in \mathbb{Z}/n\mathbb{Z}$ to multiplication by α_i (which is a K_i-linear, and hence k-linear, automorphism of K_i). Show that

$$(\rho_1, V_1), \ldots, (\rho_r, V_r)$$

are the irreducible representations of G over k, up to isomorphism.

Possible hint: See Exercise 18.3 from Lecture 18.

Note: This again illustrates that irreducible representations of an abelian group over a non-algebraically closed field need not be 1-dimensional.

20.6. **The definition of the representation ring.**

Definition 20.23. (i) Suppose $\mathcal{A}$ is an abelian category such that the isomorphism classes of finite length objects of $\mathcal{A}$ form a set. Then the Grothendieck group of $\mathcal{A}$ is the quotient of the free abelian group $F(\mathcal{A})$ of the set of isomorphism classes of finite length objects of $\mathcal{A}$, by the submodule generated by the $[V] - [V'] - [V'']$ whenever we have an exact sequence $0 \to V' \to V \to V'' \to 0$ in $\mathcal{A}$. Denote it by $K(\mathcal{A})$.

(ii) The representation ring of a finite group G over a field k is the Grothendieck group of $Rep_k(G)$, but made into a ring under the tensor product of representations (Notation 20.7(iv)). Note that the class of the trivial representation of G is the multiplicative identity of this ring. Denote the representation ring of G by $R_k(G)$.

$R_k(G)$ is generated as an abelian group by the $[V]$ as V runs over the irreducible representations of G (even in bad characteristic), and a typical element of it may be written as $\sum_i n_i V_i$ instead of as $\sum_i n_i[V_i]$.

20.7. **Induced and coinduced representations.** In this subsection, k can be any commutative ring, and will often be implicitly assumed to be fixed. We have already defined induced and coinduced representations in the context of subgroups, but the same definition applies to any homomorphism $H \to G$ of groups (though we will mostly only be interested in the subgroup case):

Definition 20.24. Let $H \to G$ be a homomorphism of groups.

(i) Ind_H^G is the extension of scalars along $k[H] \to k[G]$, i.e., $k[G] \otimes_{k[H]} - : Rep_k(H) \rightsquigarrow Rep_k(G)$.
(ii) coInd_H^G is the coextension of scalars along $k[H] \to k[G]$, i.e., $\mathrm{Hom}_{k[H]}(k[G], -) : Rep_k(H) \rightsquigarrow Rep_k(G)$.
(iii) $\mathrm{Res}_H^G : Rep_k(G) \rightsquigarrow Rep_k(H)$ is the restriction of scalars along $k[H] \to k[G]$.

The following proposition is an immediate application of Hom-tensor adjointness, exactly as in Lecture 8, though one of the assertions in (ii) of the proposition uses Proposition 20.4 – I leave it to you to review/work out the details:

Proposition 20.25. *(i) (Frobenius reciprocity)* $(\mathrm{Ind}_H^G, \mathrm{Res}_H^G, \mathrm{coInd}_H^G)$ *is an adjoint triple of functors between* $Rep_k(H)$ *and* $Rep_k(G)$*:* Ind_H^G *is left-adjoint to* Res_H^G*, and* coInd_H^G *is right-adjoint to* Res_H^G*.*

In fact, there are 'obvious maps' $\sigma \mapsto \mathrm{Res}_H^G \mathrm{Ind}_H^G \sigma =: \mathrm{Ind}_H^G \sigma|_H$ *and* $\mathrm{coInd}_H^G \sigma|_H \to \sigma$*, compositions with which realize the adjunction isomorphisms*

$$\mathrm{Hom}_G(\mathrm{Ind}_H^G \sigma, \tau) \to \mathrm{Hom}_H(\sigma, \tau|_H) \qquad \text{and} \qquad \mathrm{Hom}_G(\sigma, \mathrm{coInd}_H^G \tau) \to \mathrm{Hom}_H(\sigma|_H, \tau).$$

(ii) Ind_H^G *is right exact and* coInd_H^G *is left exact. Both are exact when* $k[G]$ *is projective over* $k[H]$*, which is the case when* $H \subset G$*, as well as when* k *is semisimple,* H *is finite, and* $\#H \in k^\times$*.*

Remark 20.26. When $H \hookrightarrow G$, the description of coinduction from Lecture 8 can be combined with Problem 1 from HW 4 to describe induction as well, giving the following descriptions of these functors:

(i) $\mathrm{coInd}_H^G(\rho, W)$ is the right-regular action of G on the space

$$\{f : G \to W \mid f(hx) = hf(x) \,\forall\, h \in H, g \in G\}.$$

(ii) $\mathrm{Ind}_H^G(\rho, W)$ is the right-regular action of G on the space

$\{f : G \to W \mid f(hx) = hf(x) \,\forall\, h \in H, g \in G,\ f$ is finitely supported on finitely many H-cosets$\}$.[62]

[62] the way induction and coinduction are defined, it only makes sense to impose the support condition on the set of H-cosets.

From the perspective of this description, the map $\mathrm{Ind}_H^G W \hookrightarrow \mathrm{coInd}_H^G W$ has the obvious description, and is immediately seen to be functorial in W. This also makes obvious the natural isomorphism, when $H \subset G$ and $[G : H]$ is finite, of the functors Ind_H^G and coInd_H^G.

Exercise 20.27. From the perspective of the above description, show that the map $W \to \mathrm{Ind}_H^G W$ (used in describing the adjunction between Ind_H^G and Res_H^G) sends $w \in W$ to the map $f : G \to W$ that is supported on H and sends $h \in H$ to $h \cdot w$. Further, the map $\mathrm{coInd}_H^G W \to W$ simply sends f to $f(1)$.

Exercise 20.28. Let $H \subset G$. Prove the following alternative description for $\mathrm{Ind}_H^G(\rho, W)$ stated in Lecture 8. Namely, $\mathrm{Ind}_H^G(\rho, W)$ is also a representation (π, V) of G, uniquely characterized up to isomorphism by the following properties:

(i) $(\pi, V)|_H$ contains a subspace isomorphic to (ρ, W); and
(ii) V is the direct sum of the translates $g_i W$ of W, as $\{g_i\}_i$ ranges over any set of representatives for G/H.

More precisely, given such a (π, V), $(\rho, W) \hookrightarrow (\pi, V)|_H$ gives via Frobenius reciprocity a unique isomorphism $\mathrm{Ind}_H^G(\rho, W) \to (\pi, V)$, whose composite with $(\rho, W) \to \mathrm{Ind}_H^G(\rho, W)|_H$ equals $(\rho, W) \hookrightarrow (\pi, V)|_H$.

Exercise 20.29. Show that whenever $H \subset G$ and $[G : H]$ is finite, $\mathrm{Ind}_H^G, \mathrm{coInd}_H^G$ define group homomorphisms:

$$\mathrm{Ind}_H^G : R_k(H) \to R_k(G), \qquad \mathrm{coInd}_H^G : R_k(H) \to R_k(G).$$

The point is that in these cases, finite length representations are taken to finite length representations. But these are not ring homomorphisms, as they don't send the multiplicativce identity to the multiplicative identity, much less do they respect the tensor product.

Proposition 20.30. *Let $H \subset G$ be a subgroup of finite index. For all representations W of H and V of G over k, we have an isomorphism, in fact a functorial one, of G-representations*

$$\mathrm{Ind}_H^G(W \otimes_k \mathrm{Res}_H^G V) \cong \mathrm{Ind}_H^G(W) \otimes_k V.$$

Proof. From $W \hookrightarrow \mathrm{Ind}_H^G(W)$, we get an injection $W \otimes_k V \hookrightarrow \mathrm{Ind}_H^G(W) \otimes_k V$ of $k[H]$-modules: although k is not a field, the injection $W \hookrightarrow \mathrm{Ind}_H^G(W)$ of k-modules is split, and hence tensoring over k preserves injectivity.

Thus, by Exercise 20.28, it is enough to show that $\mathrm{Ind}_H^G(W) \otimes_k V$ is the direct sum of $\{g_i\}_i$-translates of $W \otimes_k V$, as $\{g_i\}_i$ ranges over a set of representatives for G/H. This immediately follows from the fact that $\mathrm{Ind}_H^G(W)$ is the direct sum of $\{g_i\}_i$ translates of W. □

Corollary 20.31. *For $H \subset G$ of finite index, and k a field (so that we have defined $R_k(H)$ and $R_k(G)$ as rings) the image of the map $R_k(H) \to R_k(G)$ induced by Ind_H^G (see Exercise 20.29) is an ideal of $R_k(G)$.*

Proof. This is immediate from Proposition 20.30. □

According to Professor Nair's notes, Proposition 20.30 is an additional property of the adjunction between Ind_H^G and Res_H^G.

Exercise 20.32. Make the isomorphism $\mathrm{Ind}_H^G(W \otimes_k \mathrm{Res}_H^G V) \cong \mathrm{Ind}_H^G(W) \otimes_k V$ of Proposition 20.30 explicit, in terms of the description of induced representations given in Remark 20.26.

21. Lecture 21 – Representation theory of finite groups – II

21.1. **Mackey's formula.** Assume that G is finite, and that k is a field with $\operatorname{char} k \nmid \#G$.[63] Write $\operatorname{Int} g$ for $h \mapsto ghg^{-1}$, and ${}^g\rho$ for the representation $\rho \circ \operatorname{Int} g^{-1}$ of gHg^{-1}. Note that g 'acts on the left' of the pair (H, ρ) to give the pair $(gHg^{-1}, {}^g\rho)$, which is why we write g on the top left of ρ.

Theorem 21.1 (Mackey's formula)**.** *Let $H, K \subset G$ be subgroups, and (ρ, V) a representation of H. Then*

$$\operatorname{Res}^G_K \operatorname{Ind}^G_H \rho \cong \bigoplus_{g \in [K\backslash G/H]} \operatorname{Ind}^K_{K \cap gHg^{-1}} \operatorname{Res}^{gHg^{-1}}_{K \cap gHg^{-1}} {}^g\rho,$$

where $[K\backslash G/H]$ is a set of representatives for $K\backslash G/H$.

Proof. Recall a description of the induced representation from Lecture 20 (Exercise 20.28): $\operatorname{Ind}^G_H V$ is the unique (up to an appropriately unique isomorphism) representation W of G such that $W|_H$ contains (a copy of) the representation V of H, and whose underlying vector space W is the sum of the g-translates of V as g ranges over any set $[G/H]$ of representatives for G/H: as vector spaces,

$$\operatorname{Ind}^G_H V = W = \bigoplus_{s \in [G/H]} s \cdot V.$$

To compute its restriction to K, one breaks up the sum on the right-hand side above into K-orbits, getting

$$\operatorname{Ind}^G_H V|_K = \bigoplus_{g \in [K\backslash G/H]} Kg \cdot V$$

as representations of K. Therefore, it suffices to show that, as a representation of K,

$$\operatorname{Ind}^G_H V|_K \supset Kg \cdot V \cong \operatorname{Ind}^K_{K \cap gHg^{-1}} \operatorname{Res}^{gHg^{-1}}_{K \cap gHg^{-1}} {}^g\rho.$$

Now consider $g \cdot V$. As a representation of gHg^{-1}, it is clearly isomorphic to ${}^g\rho$. Thus, $\operatorname{Res}^K_{K \cap gHg^{-1}}(Kg \cdot V)$ contains a copy of ${}^g\rho|_{K \cap gHg^{-1}}$. Moreover, it is clear that, as vector spaces, letting $[K/(K \cap gHg^{-1})]$ be a set of representatives for $K/(K \cap gHg^{-1})$, we have

$$Kg \cdot V = \bigoplus_{a \in [K/(K \cap gHg^{-1})]} a \cdot gV$$

(since gHg^{-1} is the stabilizer of $g \cdot V$ in G, $K \cap gHg^{-1}$ is the stabilizer of $g \cdot V$ in K). Therefore, by the description of induced representations described above, we have $Kg \cdot V \cong \operatorname{Ind}^K_{K \cap gHg^{-1}} \operatorname{Res}^{gHg^{-1}}_{K \cap gHg^{-1}} {}^g\rho$, as desired. □

[63] I don't see that these conditions are used anywhere, and hence they may not be needed.

21.2. **Mackey's theorem via equivariant sheaves.** We will now describe a proof of Mackey's theorem, Theorem 21.1, by means of 'equivariant sheaves'. For us, this will be an inefficient route to it, but this seems to make the proof more natural, and the equivariant sheaves we will see are a toy model for some of the objects one sees in more sophisticated mathematics. You can consider this subsection optional. For this proof I will follow some lecture notes of Dmitry Gourevitch, sometimes closely. In what follows, a 'sheaf' will refer to a 'sheaf of vector spaces over k', where k is the field that we have fixed.

Remark 21.2. For the proof that we are going to see, it will help to know that coInd_H^G can also be realized as

$$\{f : G \to V \mid f(gh) = h^{-1}f(g) \,\forall\, h \in H, g \in G\},$$

with the left-regular action of G: $(g \cdot f)(h) = f(g^{-1}h)$. Indeed, this description is obtained from the description given in Lecture 20, Remark 20.26, by applying the change of variables $g \mapsto g^{-1}$, which switches the left and right actions. A similar comment applies to $\mathrm{Ind}_H^G V$, where one considers those f that are finitely supported modulo H.

Motivation for sheaves: One source for representations of G is as follows: let X be a set on which G acts, let $k[X]$ be the space of functions $X \to k$, and let G act on $k[X]$ according to the left-regular action, $(g \cdot f)(x) = f(g^{-1}x)$.

One can think of $k[X]$ as follows: $k[X]$ identifies with the set of sections to the obvious map $\bigsqcup_{x \in X} k \to X$. If one takes instead disjoint union over vector spaces that are allowed to depend on $x \in X$, we get a sheaf:

Definition 21.3. (i) A sheaf $\mathcal{V}$ (of k-vector spaces) on a set X is the datum of a k-vector space $\mathcal{V}_x$ for each $x \in X$. It is clear how to define morphisms between sheaves on X, so we now have a category $Sh(X)$ of sheaves on X. It is an abelian category, being the category $\mathrm{Fun}(X_{disc}, Vec_k)$ of functors from X_{disc} to Vec_k, where X_{disc} is the discrete category built from X (with $\mathrm{Ob}\, X_{disc} = X$, and the only morphisms being the identity morphisms).

(ii) For $U \subset X$ and $\mathcal{V} \in \mathrm{Ob}\, Sh(X)$, define the space of sections of $\mathcal{V}$ over U to be

$$\mathcal{V}(U) := \{s : U \to \bigsqcup_{x \in U} \mathcal{V}_x \mid s(x) \in \mathcal{V}_x \,\forall\, x \in U\}.$$

Clearly, $\mathcal{V}(U)$ is a vector space that identifies with $\bigoplus_{x \in U} \mathcal{V}_x$.

The space $\mathcal{V}(X)$ of sections of $\mathcal{V}$ over X is called the space of global sections of X, and is denoted by $\Gamma(\mathcal{V})$. Similarly, we may define the spaces $\mathcal{V}_c(U)$ and $\mathcal{V}_c(X) = \Gamma_c(\mathcal{V})$ of finitely supported sections (c for 'compact').

(iii) For $\mathcal{V} \in \mathrm{Ob}\, Sh(X)$, define the total space $T(\mathcal{V})$ of $\mathcal{V}$ to be $\bigsqcup_{x \in X} \mathcal{V}_x$, so we have a map $T(\mathcal{V}) \to X$ whose fibers are the various $\mathcal{V}_x$.

(iv) Now assume that X is a G-set for a group G, and let $\mathcal{V} \in \mathrm{Ob}\, Sh(X)$. A G-equivariant structure on $\mathcal{V}$ is an action of G on $T(\mathcal{V})$ compatible with the action of G on X, and respecting the vector space structures: in other words, G acts on $T(\mathcal{V})$ in such a way that for all $g \in G$ and $x \in X$, the action of g restricts to an isomorphism of vector spaces $\mathcal{V}_x \to \mathcal{V}_{gx}$.

(v) Let X be a G-set. A G-equivariant sheaf on X is a sheaf $\mathcal{V} \in \operatorname{Ob} Sh(X)$, together with a G-equivariant structure on it. It is clear how to define morphisms of G-equivariant sheaves, so we now have a category $Sh_G(X)$ of G-equivariant sheaves on X.

Example 21.4. If $\{*\}$ is a singleton set with the only possible G-action on it, then $Sh_G(*)$ is equivalent to $Rep_k(G)$.

Remark 21.5. (i) Let X be a G-set, let $\mathcal{V} \in \operatorname{Ob} Sh_G(X)$, and let $U \subset X$. The action of $g \in G$ then defines an isomorphism $\mathcal{V}(U) \to \mathcal{V}(gU)$, mapping $s \in \mathcal{V}(U)$ to the element of $\mathcal{V}(gU)$ defined by

$$gU \overset{g^{-1}\cdot -}{\to} U \overset{s}{\to} \bigsqcup_{x\in U} \mathcal{V}_x \overset{g}{\to} \bigsqcup_{x\in U} \mathcal{V}_{gx} = \bigsqcup_{y\in gU} \mathcal{V}_y.$$

In other words, if $s : U \to T(\mathcal{V})$ is a section, then $g \cdot s : gU \to T(\mathcal{V})$ takes $y = g \cdot x \in gU$ to $g \cdot s(x) \in g \cdot \mathcal{V}_x = \mathcal{V}_{gx}$.

If U is invariant under G, this action defines a representation of G on $\mathcal{V}(gU)$. In particular:

- For each $x \in X$, $\mathcal{V}_x$ is a representation of the isotropy group $G_x := \{g \in G \mid g \cdot x = x\}$ of x in G.
- $\Gamma(X) = \mathcal{V}(X)$ and $\Gamma_c(X) = \mathcal{V}_c(X)$ are representations of G.

Exercise 21.6. Let K be a group. If $X = \bigsqcup_i X_i$ as K-sets, then show that $Sh_K(X)$ has an obvious equivalence with the product category $\prod_i Sh_K(X_i)$ (if we haven't seen what this product category means, make sense of it). Moreover, if $\mathcal{V} \in \operatorname{Ob} Sh_K(X)$ corresponds to $(\mathcal{V}_i)_i \in \operatorname{Ob} \prod_i Sh_K(X_i)$ under this equivalence, show that we have identifications

$$\Gamma(\mathcal{V}) = \prod_i \Gamma(\mathcal{V}_i), \qquad \text{and} \qquad \Gamma_c(\mathcal{V}) = \bigoplus_i \Gamma_c(\mathcal{V}_i)$$

in $Rep_k(K)$.

Exercise 21.7. (i) If $\nu : X \to Y$ is a map of sets, define what the functor $\nu_* : Sh(X) \rightsquigarrow Sh(Y)$ of pushforward with respect to ν, and the functor $\nu^* : Sh(Y) \rightsquigarrow Sh(X)$ of pullback with respect to ν, should mean. These should satisfy, for all $\mathcal{V} \in \operatorname{Ob} Sh(X)$ and $\mathcal{W} \in \operatorname{Ob} Sh(Y)$, and all subsets $U \subset Y$ and elements $x \in X$,

$$\nu_*(\mathcal{V})(U) = \mathcal{V}(\nu^{-1}(U)), \qquad \text{and} \qquad (\nu^*(\mathcal{W}))_x = \mathcal{W}_{\nu(x)}.$$

(ii) Now assume that $X \to Y$ is a map of G-sets. Show that ν_* and ν^* extend to functors $\nu_* : Sh_G(X) \rightsquigarrow Sh_G(Y)$ and $\nu^* : Sh_G(Y) \rightsquigarrow Sh_G(X)$.

(iii) (Taken from Gourevitch's notes) Show that each of the following is an equivalent way to give a G-equivariant structure on a $\mathcal{V} \in \operatorname{Ob} Sh(X)$:

(a) The datum, for any $x \in X$ and $g \in G$, of a k-linear map $t_{g,x} : \mathcal{V}_x \to \mathcal{V}_{gx}$, such that for all $x \in X$ and $g_1, g_2 \in G$, we have $t_{g_1g_2,x} = t_{g_1,g_2\cdot x} \circ t_{g_2,x}$.

(b) An isomorphism of sheaves $\alpha : a^*(\mathcal{V}) \to p^*(\mathcal{V})$, where $a : G \times X \to X$ is the action map $(g, x) \mapsto g \cdot x$ and $p : G \times X \to X$ is the projection, satisfying the

following condition:
Writing $q, b : G \times G \times X \to X$ for the morphisms defined by $q(g_1, g_2, x) = x$ and $b(g_1, g_2, x) = (g_1 g_2 \cdot x)$, the morphisms $\beta, \gamma : q^*(\mathcal{V}) \to b^*(\mathcal{V})$ in $Sh(G \times G \times X)$ induced by α (applied twice) are equal to each other.

Hint: Given a G-equivariant sheaf $\mathcal{V}$, define $t_{g,x} : \mathcal{V}_x \to \mathcal{V}_{gx}$ to be simply given by the action of G on $\mathcal{V}$. On the other hand, define $\alpha : a^*(\mathcal{V}) \to p^*(\mathcal{V})$ to be such that

$$(a^*(\mathcal{V}))_{(g,x)} = \mathcal{V}_{g\cdot x} \overset{g^{-1}\cdot}{\to} \mathcal{V}_x = (p^*(\mathcal{V}))_{(g,x)}.$$

The 'cocycle condition' $\beta = \gamma$ is just saying that $(g_1 g_2)^{-1} : \mathcal{V}_{g_1 g_2 x} \to \mathcal{V}_x$ is the composite $(\mathcal{V}_{g_2 x} \to \mathcal{V}_x) \circ (\mathcal{V}_{g_1 g_2 x} \to \mathcal{V}_{g_2 x})$.

Note: The point is that on more sophisticated mathematical objects, where sheaves are parameterized by more 'continuous' objects like topological spaces or algebraic varieties rather than 'discrete sets', the naive definition of a G-equivariant sheaf given in Definition 21.3 above will not work since we cannot work 'point by point', but this particular "$a^*(\mathcal{V}) \cong p^*(\mathcal{V})$" definition adapts to such more general situations.

The following lemma describes induction and coinduction of representations in terms of equivariant sheaves:

Lemma 21.8. *(i) Let $H \subset G$ be a subgroup, and view G/H as a G-set. Then the functor $Sh_G(G/H) \rightsquigarrow Rep_k(H)$ defined by $\mathcal{V} \rightsquigarrow \mathcal{V}_{eH}$, where eH is the identity coset in G/H and $\mathcal{V}_{eH}$ gets the obvious action of the isotropy group $G_{eH} = H$ of $eH \in G/H$ in G (see Remark 21.5)* [64] *is an equivalence of categories.*

(ii) If the above functor is denoted by $V \rightsquigarrow \mathcal{V}$, then coInd_H^G identifies with the functor

$$Rep_k(H) \rightsquigarrow Sh_G(G/H) \rightsquigarrow Rep_k(G), \qquad \text{given by} \qquad V \rightsquigarrow \mathcal{V} \rightsquigarrow \Gamma(\mathcal{V}),$$

and Ind_H^G identifies with the analogously defined functor where one uses $\Gamma_c(\mathcal{V})$ in place of $\Gamma(\mathcal{V})$. Again, we refer to Remark 21.5 for how to view $\Gamma(\mathcal{V})$ and $\Gamma_c(\mathcal{V})$ as representations of G.

Sketch of the proof. Let us consider (i) first. I will only partly define a functor in the other direction, and leave it as an exercise to check that it works and is indeed a quasi-inverse.

If V is a representation of H (on a k-vector space), define $T(\mathcal{V})$ (the total space of the sheaf $\mathcal{V}$ that we will assign to V) to be the set of H-orbits $(G \times V)/H$, where H is made to act on $G \times V$ by $h \cdot (g, v) = (gh^{-1}, hv)$. Note that the projection map $G \times V \to G$ induces a map $T(\mathcal{V}) \to G/H$, and that each fiber of this map canonically has the structure of a vector space: the fiber over $gH \in G/H$ is the image of $\{g\} \times V \hookrightarrow G \times V \to (G \times V)/H$; this vector space structure is independent of the choice of g in its H-coset gH, since for each $h \in H$ we have a bijection $\{g\} \times V \to \{gh\} \times V$ given by $(g, v) \mapsto (gh, h^{-1}v)$, and since $v \mapsto h^{-1}v$ is a vector space isomorphism. Now check that this defines a G-equivariant

[64] As usual, we are leaving it implicit as to how to define this functor at the level of morphisms.

sheaf $\mathcal{V}$, that this assignment $V \rightsquigarrow \mathcal{V}$ is functorial, and that this functor is a quasi-inverse to the one given in the statement of the lemma.

For (ii), the point is to observe that sections to $(G \times V)/H \to G/H$ identify with maps $G \to V$ such that $f(gh) = h^{-1}f(v)$: associate to each $s \in \Gamma(\mathcal{V})$ the unique map $f_s : G \to V$ such that for all $g \in G$, $s(gH) \in (G \times V)/H$ is the image of $(g, f_s(g))$:

$$\begin{array}{ccccc} G & \ni & g \longmapsto (g, f_s(g)) & \in & G \times V \\ \downarrow & & \downarrow \quad\quad \downarrow & & \downarrow \\ G/H & \ni & gH \longmapsto s(gH) & \in & (G \times V)/H \end{array} .$$

Check that $s \mapsto f_s$ defines isomorphisms of representations of G, $\Gamma(\mathcal{V}) \to \mathrm{coInd}_H^G(V)$ and $\Gamma_c(\mathcal{V}) \to \mathrm{Ind}_H^G(V)$. □

Remark 21.9. (i) Suppose G acts on X transitively, and that $\mathcal{V} \in \mathrm{Ob}\, Sh_G(X)$. Then it follows from Lemma 21.8 that for any $x \in X$, $\Gamma(\mathcal{V}) \in \mathrm{Ob}\, Rep_k(G)$ identifies with $\mathrm{coInd}_{G_x}^G \mathcal{V}_x$, and $\Gamma_c(\mathcal{V}) \in \mathrm{Ob}\, Rep_k(G)$ identifies with $\mathrm{Ind}_{G_x}^G \mathcal{V}_x$.

(ii) If $\mathcal{V} \in \mathrm{Ob}\, Sh_G(X)$, where X is a G-set, $g \in G$ and $x \in X$, then g defines an isomorphism $\mathcal{V}_x \to \mathcal{V}_{gx}$. $\mathcal{V}_x$ is a representation of G_x (as in Remark 21.5), say ρ_{G_x}, and $\mathcal{V}_{gx}$ is a representation of $G_{gx} = gG_xg^{-1}$, say $\rho_{G_{gx}}$. Show as an exercise that $g : \mathcal{V}_x \to \mathcal{V}_{gx}$ relates the representations $(\rho_{gx}, \mathcal{V}_{gx})$ of $G_{gx} = gG_xg^{-1}$ and $(\rho_x, \mathcal{V}_x)$ of G_x as $\rho_{gx} \cong \rho_x \circ \mathrm{Int}\, g^{-1}$, i.e., $\rho_{gx} \cong {}^g\rho_x$.
More precisely, G_{gx} acts on $\mathcal{V}_{gx}$ by ρ_{gx} and on $\mathcal{V}_x$ via ${}^g\rho_x$, and the vector space isomorphism $g : \mathcal{V}_x \to \mathcal{V}_{gx}$ is also an isomorphism of representations $({}^g\rho_x, \mathcal{V}_x) \to (\rho_{gx}, \mathcal{V}_{gx})$ of the group G_{gx}.

Proof of Theorem 21.1. By Lemma 21.8, ρ corresponds to a G-equivariant sheaf $\mathcal{V}_\rho$ on $X := G/H$, and we have $\mathrm{Ind}_H^G \rho \cong \Gamma_c(\mathcal{V}_\rho)$ as a representation of G.

For $\mathrm{Res}_K^G \mathrm{Ind}_H^G \rho$, view the G-equivariant sheaf $\mathcal{V}_\rho$ as a K-equivariant sheaf (tautologically), and compute the K-action on $\Gamma_c(\mathcal{V}_\rho)$: the resulting representation is clearly $\mathrm{Res}_K^G \mathrm{Ind}_H^G \rho$.

Let $X = \bigsqcup_i K{\cdot}x_i$ be the decomposition of X into K-orbits. Thus, we can write $x_i = g_i{\cdot}(eH)$ for each i, and then $\{g_i\}_i$ is a set of representatives for $K\backslash G/H$. Accordingly, we can write $\mathcal{V}_\rho = (\mathcal{V}_{\rho,i})_i$ as in Exercise 21.6 (the $\mathcal{V}_{\rho,i}$ are K-equivariant sheaves, though of course not G-equivariant). Thus, in $Sh_K(X)$, we have by Exercise 21.6,

$$\Gamma_c(\mathcal{V}_\rho) \cong \bigoplus_i \Gamma_c(\mathcal{V}_{\rho,i}),$$

as representations of K. By Remark 21.9(i), the K-representation $\Gamma_c(\mathcal{V}_{\rho,i})$ identifies with $\mathrm{Ind}_{K_{x_i}}^K \mathcal{V}_{\rho,i,x_i}$. Since $x_i = gx_i$, we have $K_{x_i} = K \cap g_iHg_i^{-1}$. Moreover, by Remark 21.9(ii), $\mathcal{V}_{\rho,i,x_i} = \mathcal{V}_{\rho,g_i\cdot(eH)}$ identifies with ${}^{g_i}\mathcal{V}_{\rho,eH}$ as a representation of $g_iHg_i^{-1}$. Thus, $\Gamma_c(\mathcal{V}_{\rho,i})$ identifies with $\mathrm{Ind}_{K\cap g_iHg_i^{-1}}^K \mathrm{Res}_{K\cap g_iHg_i^{-1}}^{g_iHg_i^{-1}} {}^{g_i}\mathcal{V}_{\rho,eH} = \mathrm{Ind}_{K\cap g_iHg_i^{-1}}^K \mathrm{Res}_{K\cap g_iHg_i^{-1}}^{g_iHg_i^{-1}} {}^{g_i}\rho$, as desired. □

21.3. **Mackey's criterion for irreducibility.** This is a criterion for answering: when is $\operatorname{Ind}_H^G V$ irreducible?

Theorem 21.10 (Mackey's criterion)**.** *Let $k = \bar{k}$ be an algebraically closed field, and let G be a finite group such that $\operatorname{char} k \nmid \#G$. Let $H \subset G$ be a subgroup, and (ρ, V) a representation of H. For all $g \in G/H$, let ${}^gH = gHg^{-1} \cap H$ (strange notation alert). Then $\operatorname{Ind}_H^G \rho = \operatorname{Ind}_H^G V$ is irreducible if and only if both the following hold:*

(i) V is irreducible; and
(ii) For all $g \in G/H$, ${}^g\rho|_{{}^gH} = \operatorname{Res}^{gHg^{-1}}_{{}^gH}(\rho \circ \operatorname{Int} g^{-1})$ and $\operatorname{Res}^H_{{}^gH} \rho$ have no irreducible ${}^gH$-summand in common (up to isomorphism).

Proof. It is immediate that V needs to be irreducible for $\operatorname{Ind}_H^G V$ to be irreducible, so we assume without loss of generality that V is irreducible.

Since k is algebraically closed and $\operatorname{char} k \nmid \#G$ (so that $k[G]$ is semisimple), $\operatorname{Ind}_H^G V$ is irreducible if and only if $\dim_k \operatorname{End}_{k[G]}(\operatorname{Ind}_H^G \rho) = k$.

We have, using Mackey's formula once (in the second step below) and Frobenius reciprocity twice (for induction in the first step and for coinduction in the third):

$$\begin{aligned}
\operatorname{End}_G(\operatorname{Ind}_H^G \rho, \operatorname{Ind}_H^G \rho) &\cong \operatorname{Hom}_H(\rho, \operatorname{Res}_H^G \operatorname{Ind}_H^G \rho) \\
&\cong \operatorname{Hom}_H\Big(\rho, \bigoplus_{g \in [H\backslash G/H]} \operatorname{Ind}_{{}^gH}^H {}^g\rho|_{{}^gH}\Big) \\
&\cong \bigoplus_{g \in [H\backslash G/H]} \operatorname{Hom}_{{}^gH}(\rho|_{{}^gH}, {}^g\rho|_{{}^gH}),
\end{aligned}$$

where the last isomorphism uses that Ind_H^G and coInd_H^G identify with each other as functors, since $[G : H]$ is finite.

Of the above sum, the term corresponding to the identity coset $HeH = H$ is $\operatorname{Hom}_H(\rho, \rho)$, which is one-dimensional by the irreducibility of ρ. Therefore, $\operatorname{Ind}_H^G \rho$ is irreducible if and only if for all $g \in G\backslash H$, $\operatorname{Hom}_{{}^gH}(\rho|_{{}^gH}, {}^g\rho|_{{}^gH}) = 0$, which (by semisimplicity) is to say, $\rho|_{{}^gH}$ and ${}^g\rho|_{{}^gH}$ have no factors in common. □

Remark 21.11. Since this proof computes the dimension over k of the endomorphism algebra of the induced representation, it also gives us more information about the number of irreducible components (though we might not be able to calculate the exact number). There is a variant formula that helps us calculate the exact number: there is a formula for the endomorphism algebra rather than just for its dimension as a vector space over k:

$$\operatorname{End}_G(\operatorname{Ind}_H^G(\rho, V)) \cong \{f : G \to \operatorname{End}_k(V) \mid \forall\, h_1, h_2 \in H \text{ and } g \in G, f(h_1 g h_2) = \rho(h_1) f(g) \rho(h_2)\},$$

where the right-hand side is viewed as an algebra under convolution (with respect to a suitable multiple of the counting measure).

If $\pi_1 = \mathrm{Ind}_{H_1}^G(\rho_1, V_1)$ and $\pi_2 = \mathrm{Ind}_{H_2}^G(\rho_2, V_2)$, one can get a vector space isomorphism

$$\mathrm{Hom}_G(\mathrm{Ind}_{H_1}^G \rho_1, \mathrm{Ind}_{H_2}^G \rho_2) \cong$$
$$\{f : G \to \mathrm{Hom}_k(V_1, V_2) \mid f(h_2 g h_1) = \rho_2(h_2) f(g) \rho_1(h_1) \,\forall\, g \in G, h_1 \in H_1, \text{ and } h_2 \in H_2\}.$$

We have stated these formulas without proof, though the proof is not difficult. If you would like to see a proof but don't want to work it out, see, e.g., Amritanshu Prasad's notes at http://www.imsc.res.in/∼amri/html_notes/notesch1.html#x4-70001.4 .

Example 21.12. Assume that G is finite and that $(\mathrm{char}\, k, \#G) = 1$.

(i) It follows from Theorem 21.10 that if H is a normal subgroup of G, then $\mathrm{Ind}_H^G \rho$ is irreducible if and only if ρ is irreducible and $\rho \circ \mathrm{Int}\, g \ncong \rho$ for all $g \in G \backslash H$. Thus, if $D_{2n} = C_n \rtimes \mathbb{Z}/2\mathbb{Z}$ is a dihedral group of order $2n$, with C_n a cyclic group of order n, and $\chi : C_n \to \mathbb{C}^\times$ is a character, i.e., a homomorphism viewed as a one-dimensional representation, then it follows that the two-dimensional representation $\mathrm{Ind}_{C_n}^{D_{2n}} \chi$ is irreducible if χ^2 is nontrivial, and reducible if χ^2 is trivial.

(ii) If H is a proper subgroup of G, it follows from Theorem 21.10 that $\mathrm{Ind}_H^G \mathbb{1}$ is never irreducible, where $\mathbb{1}$ is the trivial representation of H over k.

(iii) If $G = H \ltimes A$ for subgroups H and A of G, with $A \subset G$ an abelian normal subgroup, the Mackey criterion allows us to describe all the irreducible complex representations of G using irreducible representations of A and irreducible representations of various subgroups of H; see HW 10 for more details.

21.4. **Representations of a product of groups.** Here I will follow Professor Nair's notes. All representations in this subsection will be over a fixed field k.

If V_1, V_2 are representations of groups G_1, G_2 over k, $V_1 \otimes_k V_2$ can be viewed as a representation of $G_1 \times G_2$: using the universal property of the tensor product, show that there exists a unique representation of $G_1 \times G_2$ on $V_1 \otimes_k V_2$ which satisfies the property that $(g_1, g_2) \cdot (v_1 \otimes v_2) = g v_1 \otimes g v_2$.

How does this generalize to the level of algebras? If R_1 and R_2 are k-algebras, and $\rho_1 : R_1 \to \mathrm{End}_k(V_1)$ and $\rho_2 : R_2 \to \mathrm{End}_k(V_2)$ are homomorphisms, the universal property of the tensor product of algebras (Exercise 19.10 from Lecture 19) gives us a homomorphism:

$$\rho = \rho_1 \otimes_k \rho_2 : R_1 \otimes_k R_2 \to \mathrm{End}_k(V_1) \otimes_k \mathrm{End}_k(V_2) \to \mathrm{End}_k(V_1 \otimes_k V_2)$$

(the latter map is just from the functoriality of the tensor product, as observed in HW 3).

To relate this to the tensor product of representations of groups, note that $k[G_1] \hookrightarrow k[G_1 \times G_2]$ and $k[G_2] \hookrightarrow k[G_1 \times G_2]$, obtained from the inclusions $G_1 \cong G_1 \times \{1\} \hookrightarrow G_1 \times G_2$ and $G_2 \cong \{1\} \times G_2 \hookrightarrow G_1 \times G_2$, give us by the universal property of tensor products of algebras a homomorphism $k[G_1] \otimes_k k[G_2] \to k[G_1 \times G_2]$, which is readily checked to be an isomorphism: both sides have k-bases indexed by $G_1 \times G_2$, which the above map respects.

It is then easy to see that the identification of $Rep_k(G)$ with $k[G]$-Mod for $G = G_1, G_2, G_1 \times G_2$ 'transport the definition of tensor product for representations of groups to the definition of tensor product for representations of algebras'.

Remark 21.13. Earlier, we defined a slightly different notion of tensor product for representations: if (ρ_1, V_1) and (ρ_2, V_2) are representations of G, we defined the representation $(\rho_1 \otimes_k \rho_2, V_1 \otimes_k V_2)$ of G (see Lecture 20, Notation 20.7**??**). This 'internal' tensor product is related to the above 'external' tensor product as follows: if you form the 'external' tensor product $\rho_1 \otimes_k \rho_2 : G \times G \to GL_k(V_1 \otimes_k V_2)$ of $G \times G$ as defined in the above discussion, then composing it with the diagonal map $\Delta : G \to G \times G$ given by $g \mapsto (g, g)$ gives the 'internal' tensor product $G \to GL(V_1 \otimes_k V_2)$.

Proposition 21.14. *Let G_1, G_2 be finite groups, and $k = \bar{k}$ an algebraically closed field.*

(i) If V_i is an irreducible representation of G_i for $i = 1, 2$, then $V_1 \otimes_k V_2$ is an irreducible representation of $G_1 \times G_2$.

(ii) Any irreducible representation of $G_1 \times G_2$ is isomorphic to $V_1 \otimes_k V_2$, where V_1 and V_2 are respectively irreducible representations of G_1 and G_2.

Proof. We will identify $k[G_1 \times G_2]$ with $k[G_1] \otimes_k k[G_2]$ as explained in the discussion preceding the proposition.

Let us prove (i). By Burnside's theorem (which uses that k is algebraically closed), $k[G_1] \to \mathrm{End}_k(V_1)$ and $k[G_2] \to \mathrm{End}_k(V_2)$ are surjective, and hence so is $k[G_1] \otimes_k k[G_2] \to \mathrm{End}_k(V_1) \otimes_k \mathrm{End}_k(V_2)$, by the right-exactness of the tensor product. Our identifications $k[G_1] \otimes_k k[G_2] \to k[G_1 \times G_2]$ and $\mathrm{End}_k(V_1) \otimes_k \mathrm{End}_k(V_2) \to \mathrm{End}_k(V_1 \otimes_k V_2)$ now give a surjective morphism $k[G_1 \times G_2] \to \mathrm{End}_k(V_1 \otimes_k V_2)$, and this morphism is clearly associated to $V_1 \otimes_k V_2$. Since $V_1 \otimes_k V_2$ is a simple module over $\mathrm{End}_k(V_1 \otimes_k V_2)$, this gives (i).

We now come to (ii). Let R_1, R_2 be finite dimensional k-algebras. Assume that we are given an irreducible left R-module V, where $R = R_1 \otimes_k R_2$. For $i = 1, 2$, we will view a left $R_1 \otimes_k R_2$-module also as a left R_i-module, via $R_i \hookrightarrow R_1 \otimes_k R_2$. [65] It is enough to show that there exist an irreducible left R_1-module V_1 and an irreducible left R_2-module V_2 such that $V \cong V_1 \otimes_k V_2$ as $R_1 \otimes_k R_2$-modules.

First assume that R_1 and R_2 are semisimple. Let V_1 be an irreducible left R_1-module contained in $V|_{R_1}$. Since V is an irreducible left $R_1 \otimes_k R_2$-module, V is the sum of R_2-translates of V_1. Each such R_2-translate is a left R_1-module that is a homomorphic image of the simple R_1-module V_1, and is hence isomorphic to either 0 or V_1. Therefore, $V|_{R_1}$ breaks up as a direct sum of copies of V_1 (this uses the semisimplicity of R_1). Similarly, $V|_{R_2}$ breaks up as a direct sum of copies of a simple left R_2-module V_2. This implies that $R_1 \to \mathrm{End}_k(V)$ and $R_2 \to \mathrm{End}_k(V)$ factor through $R_1 \twoheadrightarrow \mathrm{End}_k(V_1)$ and $R_2 \twoheadrightarrow \mathrm{End}_k(V_2)$ (the two '$\twoheadrightarrow$'s are by Burnside's theorem, because V_i is a simple left R_i-module for $i = 1, 2$). Therefore, $R_1 \otimes_k R_2 \to \mathrm{End}_k(V)$ is trivial on the kernel of $R_1 \otimes_k R_2 \to \mathrm{End}_k(V_1) \otimes_k \mathrm{End}_k(V_2) \cong \mathrm{End}_k(V_1 \otimes_k V_2)$. [66] Because $R_1 \otimes_k R_2 \to \mathrm{End}_k(V_1) \otimes_k \mathrm{End}_k(V_2) \cong \mathrm{End}_k(V_1 \otimes_k V_2)$ is a surjection, $R_1 \otimes_k R_2 \to \mathrm{End}_k(V)$ factors through it. Thus, the left R-module structure on V is obtained by pulling back a left $\mathrm{End}_k(V_1 \otimes_k V_2)$-module structure on V under

[65] This is injective since anything over a field is flat.

[66] Check using vector space bases that this kernel is just $\ker(R_1 \to \mathrm{End}_k(V_1)) \otimes R_2 + R_1 \otimes \ker(R_2 \to \mathrm{End}_k(V_2)))$.

$R_1 \otimes_k R_2 \to \mathrm{End}_k(V_1) \otimes_k \mathrm{End}_k(V_2) \cong \mathrm{End}_k(V_1 \otimes_k V_2)$. Since $\mathrm{End}_k(V_1 \otimes_k V_2)$ has a unique simple left module up to isomorphism, namely $V_1 \otimes_k V_2$, it follows that $V \cong V_1 \otimes_k V_2$ as a module over $\mathrm{End}_k(V_1 \otimes_k V_2)$, and hence as a module over $R = R_1 \otimes_k R_2$.

Now we come to the general case, where R_1 and R_2 are not semisimple. Let $\bar{R}_1 = R_1/\operatorname{rad}(R_1)$ and $\bar{R}_2 = R_2/\operatorname{rad}(R_2)$. Since $k = \bar{k}$, it is easy to see that $\bar{R}_1 \otimes_k \bar{R}_2$ is a semisimple k-algebra: by the Artin-Wedderburn theorem, and the fact that there is no finite dimensional division algebra over k (since k is algebraically closed), this reduces to showing that each $M_{n_1}(k) \otimes_k M_{n_2}(k)$ is a semisimple k-algebra, which it is, being isomorphic to $M_{n_1 n_2}(k)$.

We claim that $R_1 \otimes_k R_2 \to \bar{R}_1 \otimes_k \bar{R}_2$ is the maximal semisimple quotient of $R_1 \otimes_k R_2$, i.e., $\bar{R}_1 \otimes_k \bar{R}_2 \cong (R_1 \otimes_k R_2)/\operatorname{rad}(R_1 \otimes_k R_2)$. Since $\bar{R}_1 \otimes_k \bar{R}_2$ is semisimple, this follows if we show that

$$\ker(R_1 \otimes_k R_2 \to \bar{R}_1 \otimes_k \bar{R}_2) = \operatorname{rad}(R_1) \otimes_k R_2 + R_1 \otimes_k \operatorname{rad}(R_2)$$

is contained in $\operatorname{rad}(R_1 \otimes_k R_2)$. This in turn follows from the fact that $\operatorname{rad}(R_1)$ and $\operatorname{rad}(R_2)$ are nilpotent (which they are, R_1 and R_2 being Artin rings – see Lemma 17.10 from Lecture 17), so that $\operatorname{rad}(R_1) \otimes_k R_2 + R_1 \otimes_k \operatorname{rad}(R_2)$ is nilpotent as well, and the fact that nilpotent left ideals in an Artin ring are contained in its Jacobson radical (see Lemma 17.14 from Lecture 17).

This proves the claim that $R_1 \otimes_k R_2 \to \bar{R}_1 \otimes_k \bar{R}_2$ is a maximal semisimple quotient. Hence irreducible representations of $R_1 \otimes_k R_2, R_1$ and R_2 identify respectively with those of $\bar{R}_1 \otimes_k \bar{R}_2, \bar{R}_1$ and $\bar{R}_2$, and we are reduced to the case where R_1 and R_2 are semisimple, which has already been taken care of. □

Remark 21.15. (i) In the above proposition, the condition that k is algebraically closed cannot be dropped: show that the 'rotation by 90°' action of $\mathbb{Z}/4\mathbb{Z}$ on $V = \mathbb{R}^2$ satisfies that V is an irreducible representation of $\mathbb{Z}/4\mathbb{Z}$, but that $V \otimes_k V$ is a reducible representation of $\mathbb{Z}/4\mathbb{Z}$.
Hint: Make use of the fact that $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \cong \mathbb{C} \times \mathbb{C}$.

(ii) Here is a more 'abstract' reason why k being algebraically closed cannot be dropped. If V is an irreducible representation of G over k, then it is easy to check (from semisimplicity) that so is $V^\vee$, so it suffices to explain why $\mathrm{End}_k(V) = V^\vee \otimes_k V$ is not in general semisimple for the action of $G \times G$ on $\mathrm{End}_k(V)$ transferred from $V^\vee \otimes_k V$ (work this action out as an easy exercise). But this is because if $D = \mathrm{End}_G(V)$ is the associated division algebra, then $\mathrm{End}_D(V) \subset \mathrm{End}_k(V)$ is a $G \times G$-invariant subspace that is proper whenever $D \neq k$.

21.5. **The Schur orthogonality relations – the abelian case.** In this subsection, we will deal with a finite group G, and a field k. Eventually we will impose two conditions: the good characteristic condition that $(\#G, \operatorname{char} k) = 1$, and also that $k = \bar{k}$ is algebraically closed.

Recall that $k[G]$ can be thought of as the ring of formal linear combinations $\sum_{g \in G} a_g g$ with each $a_g \in k$, and also as the space of functions $G \to k$ under convolution. Today,

we will take the latter perspective on $k[G]$, and view it as also a representation of $G \times G$: $(g_1, g_2) \cdot f(g) = f(g_1^{-1} g g_2)$: the first copy of G acts by the left-regular representation, and the second copy of G acts by the right-regular representation. This representation of $G \times G$ on $k[G]$ is called the regular representation.

Let $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ be the irreducible representations of G over k up to isomorphism. Recall that the action map

$$k[G] \to \prod_{i=1}^{r} \mathrm{End}_k(V_i) \tag{80}$$

quotients to an isomorphism of rings

$$k[G]/(\mathrm{rad}(k[G])) \to \prod_{i=1}^{r} \mathrm{End}_{D_i}(V_i), \tag{81}$$

where $D_i = \mathrm{End}_G(V_i)$ is the division algebra associated to the representation ρ_i.

(81) is a form of the Fourier transform, in good characteristic (see the discussion on $L^2(S^1)$ below). It can be written as

$$(f : G \to k) \mapsto \Big(\sum_{g \in G} f(g) \rho_i(g) \Big)_i = \Big(\int_{g \in G} f(g) \rho_i(g) \, dg \Big)_i. \tag{82}$$

The fact that (81) converts convolution in $k[G]$ to multiplication in $\prod_{i=1}^{r} \mathrm{End}_{D_i}(V_i)$ is a general property of Fourier transforms, which we see in the classical situations as well.

Note that in good characteristic and when k is algebraically closed, (80) itself is an isomorphism of rings. Therefore, we would like to invert it explicitly.

In the rest of this lecture, we will work out a formula for this 'Fourier inversion' in the special case where G is abelian as well, through a series of exercises.

Before the series of exercises, let us describe its setting. Assume that G is finite abelian, $(\#G, \mathrm{char}\, k) = 1$, and that $k = \bar{k}$ is algebraically closed. Recall that in this case, the set of irreducible representations of G on k-vector spaces, up to isomorphism, can be identified with $\mathrm{Hom}_{k\text{-}Alg}(k[G], k)$, which further by restriction along $G \hookrightarrow k[G]^\times$ identifies with $\mathrm{Hom}(G, k^\times)$.

Let us write (81) as:

$$f \mapsto \Big(\sum_{g \in G} f(g) \chi(g) \Big)_\chi = \Big(\int_{g \in G} f(g) \chi(g) \, dg \Big)_\chi. \tag{83}$$

Remark 21.16. When G is abelian, k is algebraically closed and $(\#G, k^\times) = 1$, it is an easy exercise to see that $\hat{G} := \mathrm{Hom}(G, k^\times)$ has the same cardinality as G (e.g., use the structure theorem for abelian groups). Thus, this abelian (good characteristic algebraically closed) case has the particular property that the set $Irr(G)$ of irreducible representations of G up to isomorphism, itself has the structure of a group, namely $\hat{G} = \mathrm{Hom}(G, k^\times)$. Note

that we then have an obvious map $G \to \hat{\hat{G}}$, which is readily verified to be an isomorphism. This is one of the most basic variants of Pontrjagin duality.

Exercise 21.17. (Simple once you do and absorb it, and very important). Assume that G is abelian, k is algebraically closed, and $(\operatorname{char} k, \#G) = 1$.

(i) Prove the following 'Schur orthogonality relations':

(a) Schur orthogonality for characters: If $\chi_1, \chi_2 \in \hat{G} = \operatorname{Hom}(G, k^\times)$ then

$$\frac{1}{\#G} \sum_{g \in G} \chi_1(g^{-1})\chi_2(g) = \begin{cases} 1, & \text{if } \chi_1 = \chi_2, \text{ and} \\ 0, & \text{otherwise.} \end{cases} \tag{84}$$

In other words, we have found an inverse image to the Dirac delta at each $\chi \in \operatorname{Hom}(G, k^\times)$ under (83): it is the map $G \to k$ given by:

$$\frac{1}{\#G} \sum_{g \in G} \chi(g^{-1})g.$$

(b) Schur orthogonality for conjugacy classes: if $a, b \in G$, then

$$\frac{1}{\#G} \sum_{\chi \in \operatorname{Hom}(G,k^\times)} \chi^{-1}(b)\chi(a) = \begin{cases} 1, & \text{if } a = b, \text{ and} \\ 0, & \text{otherwise.} \end{cases} \tag{85}$$

This gives an explicit description of the Dirac delta function at $b \in G$ in $k[G]$ as a linear combination of the $\chi \in \operatorname{Hom}(G, k^\times) \subset k[G]$: namely,

$$\frac{1}{\#G} \sum_{\chi \in \operatorname{Hom}(G,k^\times)} \chi^{-1}(b)\chi \in k[G].$$

(ii) Using these relations or otherwise, prove that the isomorphism

$$k[G] \to \prod_{\chi \in \operatorname{Hom}(G,k^\times)} k$$

from (83) has the following inverse:

$$c = (\chi \mapsto c(\chi))_{\chi \in \operatorname{Hom}(G,k^\times)} \mapsto (\#G)^{-1} \cdot \sum_{\chi \in \operatorname{Hom}(G,k^\times)} c(\chi) \cdot \chi^{-1} = \int_{\chi \in \operatorname{Hom}(G,k^\times)} c(\chi) \cdot \chi^{-1}\, d\chi \tag{86}$$

(note that $\chi^{-1} \in \operatorname{Hom}(G, k^\times) \subset k[G]$ for each χ), where $d\chi$ is $(\#G)^{-1}$ times the counting measure on the 'Pontrjagin dual' group $\hat{G} = \operatorname{Hom}(G, k^\times)$. This measure can be thought of as dual to the counting measure on G.

(iii) Show that the isomorphism

$$k[G] \to \bigoplus_{\chi \in \operatorname{Hom}(G,k^\times)} k,$$

where we now write $\bigoplus$ instead of $\prod$ as we wish to ignore the ring structures, transports the symmetric nondegenerate bilinear form

$$\langle f_1, f_2 \rangle = \frac{1}{\#G} \sum_{g \in G} f_1(g^{-1}) f_2(g)$$

on the left-hand side to the symmetric nondegenerate bilinear form

$$\langle c_1, c_2 \rangle = \sum_{\chi \in \mathrm{Hom}(G, \mathbb{C}^\times)} c_1(\chi) c_2(\chi).$$

(iv) When $k = \mathbb{C}$ in addition, state the Schur orthogonality relations for characters can be stated in terms of an inner product, and prove them.

(v) Read and make sense of the following statements. $k \cdot \chi \subset k[G]$ is the (χ^{-1}, χ)-isotypic component of the regular representation of $G \times G$ on $k[G]$, and thus

$$k[G] \cong \bigoplus_{\chi \in \mathrm{Hom}(G, k^\times)} k \cdot \chi$$

is the decomposition of $k[G]$ into isotypic subspaces for the regular representation. Note that this takes the form

$$(k[G], \text{regular representation}) \cong \bigoplus_{(\rho, V)} V^\vee \otimes_k V,$$

where (ρ, V) ranges over the set of irreducible representations of G up to isomorphism. Relatedly, the above decomposition is the simultaneous diagonalization of the regular action of $G \times G$ on $k[G]$. Similar assertions apply with the regular representation replaced by the left regular and the right regular representations. Thus, decomposition into isotypic components is (in some ways) a generalization of simultaneous diagonalization, and irreducible representations are like 'families of eigenvalues'.

(vi) (Fourier expansion) Make sense of the following related point as well: since $\mathrm{Hom}(G, k^\times) \subset k[G]$ is a particularly nice 'eigen' basis, we might like to express a function $f \in k[G]$ in terms of this 'eigen' basis $\mathrm{Hom}(G, k^\times)$. Together, (83) and (86) let us expand each $f \in k[G]$ as

$$f = \sum_{\chi \in \mathrm{Hom}(G, k^\times)} a_\chi \chi,$$

where $(c_\chi)_\chi$ is the image of f under (83), and (by (86)) $a_\chi = (\#G)^{-1} c_{\chi^{-1}}$. Thus, Fourier expansion is analogous to expressing a vector in terms of an eigenbasis for a family of operators.

Example 21.18. (i) The considerations of Exercise 21.17 can be adapted to compact abelian Lie groups, as we outline in special cases without any formal justifications. Consider the compact but infinite group $G = S^1$. In this case, one analogue of $k[G]$ is $L^2(S^1)$, whose elements can be thought of as periodic functions on $\mathbb{R}$ with period 1. In this case, the analogue of $\hat{G} = \mathrm{Hom}(G, k^\times)$ is $Hom_{cts}(S^1, \mathbb{C}^\times) = Hom_{cts}(S^1, S^1)$, which can be identified with $\mathbb{Z}$ (thought of as a discrete topological

group): $n \in \mathbb{Z}$ identifies with $\chi_n : z \mapsto z^n$ on S^1, which corresponds to the periodic function $x \mapsto e^{2\pi i n x}$. The analogue of (83) is:

$$f \mapsto \Big(\hat{f} : n \mapsto \int_{S^1} f(z)\chi_n(z)\,dz\Big) = \int_0^1 ((f \circ e^{2\pi i})(x))e^{2\pi i n x}\,dx.\ ^{67}$$

The analogue of (86) is:

$$g \mapsto \Big(z \mapsto \sum_{n\in\mathbb{Z}} g(-n)z^n\Big)$$

(interpret it at the level of periodic functions for a more classical formulation).

(ii) Suppose $G = \mathbb{R}, k = \mathbb{C}$. In this case, again we consider $L^2(G)$ in place of $k[G]$. The analogue of $\hat{G}$ that is relevant to us here turns out to be not $Hom_{cts}(\mathbb{R}, \mathbb{C}^\times)$ but:

$$Hom_{cts}(\mathbb{R}, S^1) = \{\exp(iy-) \mid y \in \mathbb{R}\}.$$

The analogue of (83) is:

$$f \mapsto \Big(\hat{f} : x \mapsto \int_{\mathbb{R}} f(x)e^{ixy}\,dy\Big),$$

and the analogue of (86) is:

$$g \mapsto \frac{1}{2\pi}\Big(x \mapsto \int_{\mathbb{R}} g(y)e^{-iyx}\,dy\Big).$$

Now we discuss two examples of applications of Fourier inversion:

Example 21.19. (i) Let $k = \mathbb{C}, G = \mathbb{Z}/a\mathbb{Z}$. Let us recall why $\sum_p (1/p)$, where p ranges over the prime numbers, diverges. In this example, any sum or product over 'n' will be over the positive integers, and any sum or product over 'p' will be over the prime numbers. Using the product expansion

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

for $\mathrm{Re}(s) > 1$, and the expansion for $\log(1+x)$, it is easy to see that as $s \to 1+$,

$$\log \zeta(s) = \sum_p \frac{1}{p^s} + (\text{something bounded as } s \to 1+).$$

This gives that $\lim_{s\to 1+} \sum_p p^{-s} = \infty$, so $\sum_p (1/p)$ diverges. If one refines this argument, one can get the prime number theorem.

Now, we crudely outline how Dirichlet proved his famous theorem on primes in arithmetic progressions – that $\sum_{p \equiv b \pmod a} p^{-s}$, where $(a, b) = 1$, diverges at 1. Problem: considering $\sum_{n \equiv b \pmod a} n^{-s}$ does not help: it does not have a product expansion, and does not seem related to $\sum_{p \equiv b \pmod a} p^{-s}$.

[67] As one of you pointed out, many if not all classical definitions have $e^{-2\pi i n x}$ instead of $e^{2\pi i n x}$: this difference is not essential for our purposes.

Dirichlet had the idea of Fourier-inverting the situation. He noticed that:

$$\sum_{p\equiv b \pmod{a}} p^{-s} = (\#(\mathbb{Z}/a\mathbb{Z})^\times)^{-1} \cdot \sum_p \sum_{\chi\in \mathrm{Hom}(\mathbb{Z}/N\mathbb{Z},\mathbb{C}^\times)} \chi(b)^{-1} \cdot \frac{\chi(p)}{p^s}$$

(formally, this follows from (86)). Here, each $\chi : (\mathbb{Z}/a\mathbb{Z})^\times \to \mathbb{C}^\times$ is viewed as a function $\chi : \mathbb{Z} \to \mathbb{C}$ that vanishes at integers that are not relatively prime to a.

The point is that, $\sum_p \chi(p)p^{-s}$ is a good thing: because $\chi : \mathbb{Z} \to \mathbb{C}^\times$ is multiplicative, it is easy to see that on setting $L(s,\chi) = \sum_{n\geqslant 1} \chi(n)n^{-s}$ as the 'χ-analogue of $\zeta(s)$', we have that as $s \to 1+$,

$$\log L(s,\chi) = \sum_p \frac{\chi(p)}{p^s} + (\text{something bounded as } s \to 1+).$$

One can show that $L(s,\chi)$ stays away from 0 and ∞ near $s = 1$ as long as $\chi : (\mathbb{Z}/a\mathbb{Z})^\times \to \mathbb{C}^\times$ is not trivial (justification for this is one of the important technical inputs into the proof), and hence $\log L(s,\chi)$ makes sense and remains bounded near $s = 1$ for nontrivial χ. This implies that as $s \to 1+$,

$$\sum_{p\equiv b \pmod{a}} p^{-s}$$

has the same asymptotics as $(\#(\mathbb{Z}/a\mathbb{Z})^\times)^{-1} \cdot \sum_p p^{-s}$. This gives the infiniteness of such p, and also shows that as b varies over numbers relatively prime to a, they all occur with similar asymptotics to each other.

(ii) Polya's inequality; I will be extra crude here. If χ is a primitive Dirichlet character modulo k (i.e., $\chi : (\mathbb{Z}/k\mathbb{Z})^\times \to \mathbb{C}^\times$ does not factor through $(\mathbb{Z}/k\mathbb{Z})^\times \to (\mathbb{Z}/l\mathbb{Z})^\times$ for any proper factor l of k), then for all $x \geqslant 1$ we have:

$$\Big|\sum_{n\leqslant x} \chi(n)\Big| < \sqrt{k}\log k.$$

The proof involves taking the Fourier expansion of χ, viewed as a function $\mathbb{Z}/k\mathbb{Z} \to \mathbb{C}$ (note that on $\mathbb{Z}/k\mathbb{Z}$, as opposed to on $(\mathbb{Z}/k\mathbb{Z})^\times$, χ is not a homomorphism or anything). Why $\mathbb{Z}/k\mathbb{Z}$? Because, crudely speaking, χ is being summed and not multiplied; we are almost integrating it against the trivial character of $\mathbb{Z}/k\mathbb{Z}$. By Schur orthogonality, only the trivial 'Fourier coefficient' of χ contributes, and this (it turns out) reduces us to estimating the

$$\sum_{m\leqslant x} e^{-2\pi imn/k} \qquad 1 \leqslant n \leqslant k-1.$$

Let me repeat the above explanation. We are adding the $\chi(n)$, but χ is a character of $(\mathbb{Z}/k\mathbb{Z})^\times$ and not of $\mathbb{Z}/k\mathbb{Z}$. Characters ψ of $\mathbb{Z}/k\mathbb{Z}$ are the ones for which estimating sums like $\sum_{n\leqslant x} \psi(x)$ is easy, since addition is the operation in $\mathbb{Z}/k\mathbb{Z}$ unlike in $(\mathbb{Z}/k\mathbb{Z})^\times$, and hence Schur orthogonality relations for $(\mathbb{Z}/k\mathbb{Z})$ help. Thus, one expresses χ in terms of additive characters of $\mathbb{Z}/k\mathbb{Z}$, which one can by Fourier-expanding χ, viewed as a function on $\mathbb{Z}/k\mathbb{Z}$.

22. Lecture 22 – Schur orthogonality relations and Fourier inversion

We begin by setting some conventions and stating the informal goal for today's lecture. Throughout this lecture, G will be a finite group, and k will be a field such that $(\operatorname{char} k, \#G) = 1$, and any representation will be finite dimensional and over k – sometimes we will mention this hypothesis explicitly, but even if we do not do so, it will be assumed to hold. We will occasionally, but not always, assume that k is algebraically closed. If we talk of $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ without defining these, then they will be understood to be the irreducible representations of G up to isomorphism.

Recall that we have an isomorphism of rings:

$$\prod_{i=1}^{r} \rho_i : k[G] \to \prod_{i=1}^{r} \operatorname{End}_{D_i}(V_i), \tag{87}$$

where $D_i := \operatorname{End}_G(V_i) := \operatorname{End}_{k[G]}(Vi)$ is the division algebra associated to ρ_i.

Thus, one would like to invert this map.

What we will do today: When $k = \bar{k}$, we will construct an inverse explicitly, but *without using* the theorems of Artin-Wedderburn, Jacobson, Burnside etc. In particular, we will avoid making use of any prior knowledge that $\prod_{i=1}^{r} \rho_i$ is an isomorphism, but instead rederive it by constructing an inverse. To do this, we will prove and use the Schur orthogonality relations (whose abelian version was given as Exercise 21.17 in Lecture 21).

Notation 22.1. Throughout this lecture, we will view $k[G]$ as a representation of $G \times G$ via the regular representation: $(g_1, g_2) \cdot f(g) = f(g_1^{-1} g g_2)$.

22.1. **The matrix coefficient map.** We will think of the two sides of (87) as representations of $G \times G$ rather than as rings, for which we need to define how $G \times G$ acts on $\operatorname{End}_k(V_i)$. Unfortunately, there are two such actions of relevance to us, slightly different from each other:

Notation 22.2. Let (ρ, V) be a representation of G.

(i) The first action of $G \times G$ on $\operatorname{End}_k(V)$ that we will consider is given by:

$$(g_1, g_2) \cdot A = \rho(g_1) A \rho(g_2)^{-1}. \tag{88}$$

Note the relevance of this action: with this action, the map (87) is a homomorphism of representations of $G \times G$.

(ii) The second action of $G \times G$ on $\operatorname{End}_k(V)$ that we will consider is given by:

$$(g_1, g_2) \cdot A = \rho(g_2) A \rho(g_1)^{-1}. \tag{89}$$

The relevance of this is that it will make the 'matrix coefficient map' that we will define below, in the 'opposite direction', a map of representations of $G \times G$.

Note that the usual structure of a representation of G on $\mathrm{End}_k(V) = \mathrm{Hom}_k(V, V)$ is the restriction of either of the actions along the diagonal map $\Delta : G \hookrightarrow G \times G$. We will refer to these actions as the 'first action' and the 'second action' of $G \times G$ on $\mathrm{End}_k(V)$, but to avoid confusion, each time we do so we will link to the appropriate equation, (88) or (89).

Since we are thinking of (87) as a map of $G \times G$-representations rather than of rings, we replace the '$\prod$' in it with '$\bigoplus$', to write it as a map of representations of $G \times G$,

$$k[G] \to \bigoplus_{i=1}^{r} \mathrm{End}_k(V_i) \to \bigoplus_{i=1}^{r} (V_i^\vee \otimes_k V_i), \tag{90}$$

where we give each $\mathrm{End}_k(V_i)$ the first action (88) of $G \times G$, and where the isomorphism $V_i^\vee \otimes_k V_i \to \mathrm{End}_k(V_i)$ was described in HW 3 (it is partially reviewed in Notation 22.3 below). The resulting action on $V_i^\vee \otimes_k V_i$ is not the usual one, so for now we ignore it.

For now, we only assume that $(\mathrm{char}\, k, \#G) = 1$, so (90) may not be an isomorphism.

22.2. **Matrix coefficients and the matrix coefficient map.** In this section, we define a map in the opposite direction of (90) (but not quite inverse to it).

Notation 22.3. Let V be a finite dimensional vector space over k. The image of $u \otimes v \in V^\vee \otimes_k V$ under the isomorphism $V^\vee \otimes_k V \to \mathrm{End}_k(V)$ (of HW 3) will be denoted by $A_{u,v}$. Thus, for all $u \in V^\vee$ and $v \in V$, $A_{u,v} \in \mathrm{End}_k(V)$ is the rank one operator defined by:

$$A_{u,v}(v') = \langle u, v' \rangle \cdot v. \tag{91}$$

Remark 22.4. Later, we will use that $\mathrm{tr}(A_{u,v}) = \langle u, v \rangle$: this has been seen in HW 3 as the commutativity of the following diagram:

$$\begin{array}{ccc} V^\vee \otimes_k V & \xrightarrow{\cong} & \mathrm{End}_k(V) \\ & {\scriptstyle ev} \searrow \quad \swarrow {\scriptstyle \mathrm{tr}} & \\ & k & \end{array},$$

where $ev : V^\vee \otimes_k V \to k$ is the evaluation map, induced by the universal property of tensor product from the tautological bilinear pairing $V^\vee \times V \to k$.

Remark 22.5. Recall from Lecture 21 that if (ρ, V) is a representation of G, then $V^\vee \otimes V$ is a representation of $G \times G$: for all $(g_1, g_2) \in G \times G, u \in V$ and $v \in V^\vee$, we have

$$(g_1, g_2) \cdot (u \otimes v) = (g_1 \cdot u, g_2 \cdot v) = (u \circ (g_1^{-1} \cdot -), g_2 \cdot v). \tag{92}$$

Via the isomorphism $V^\vee \otimes_k V \to \mathrm{End}_k(V)$, this makes $\mathrm{End}_k(V)$ into a representation of $G \times G$ as well; since $A_{u \circ \rho(g_1^{-1}), \rho(g_2) \circ v} = \rho(g_2) \circ A_{u,v} \circ \rho(g_1)^{-1}$, this action of $G \times G$ on $\mathrm{End}_k(V)$ is the second action from Notation 22.2, i.e., given by (89), which we repeat as:

$$(g_1, g_2) \cdot A = \rho(g_2) \circ A \circ \rho(g_1^{-1}).$$

We will now define maps in the opposite direction to (90): these will not be inverses to (90) even when k is algebraically closed, but composites each way will be simple enough that we will be able to define the inverses easily.

If (ρ, V) is a representation of G, how can we define elements of $k[G]$ using it? One way is to think of $\rho : G \to \mathrm{GL}_k(V)$ as $\rho : G \to GL_n(k)$ using a basis $e_1, \ldots, e_n$ of V, and for any $1 \leqslant i, j \leqslant n$ define:

$$s_{ij}(g) = (i,j)\text{-th matrix entry of } \rho(g) = \langle e_i^\vee, \rho(g) e_j \rangle.$$

While this depends on the basis, the latter expression above tells us how to get a coordinate-free version:

Definition 22.6. (i) Let $\rho : G \to \mathrm{GL}_k(V)$ be a finite dimensional representation. If $u \in V^\vee$ and $v \in V$, the element the matrix coefficient of ρ associated to u and v is defined to be:

$$c_{u,v} : (g \mapsto \langle u, \rho(g)v \rangle) \in k[G].$$

(ii) Since $(u, v) \mapsto c_{u,v} \in k[G]$ is bilinear, we get a map

$$\mathrm{End}_k(V) = V^\vee \otimes_k V \to k[G]$$

characterized by the property that $u \otimes v \mapsto c_{u,v}$. This is called the matrix coefficient map associated to (ρ, V). In Exercise 22.7(i) below, you will verify that, as a map $\mathrm{End}_k(V) \to k[G]$, it is given by $A \mapsto (g \mapsto \mathrm{tr}(\rho(g)A))$.

(iii) This gives us our 'map in the other direction', the matrix coefficient map for G, obtained by adding up the matrix coefficient maps associated to the irreducible representations $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ of G up to isomorphism. Explicitly, it is given by:

$$\bigoplus_{i=1}^{r} \mathrm{End}_k(V_i) \to k[G], \qquad (A_i)_i \mapsto \Big(g \mapsto \sum_{i=1}^{r} \mathrm{tr}(\rho_i(g) A_i) \Big). \tag{93}$$

Exercise 22.7. Let (ρ, V) be a finite dimensional representation of G.

(i) Show that in terms of $\mathrm{End}_k(V)$ as opposed to $V^\vee \otimes_k V$, the matrix coefficient map $\mathrm{End}_k(V) \to k[G]$ is given by $A \mapsto (g \mapsto \mathrm{tr}(\rho(g)A))$.

(ii) Verify that the matrix coefficient map $V^\vee \otimes_k V \to k[G]$ is a map of $G \times G$-representations, where $G \times G$ acts on $V^\vee \otimes_k V$ as in Remark 22.5, and on $k[G]$ of course by the regular representation.

(iii) Conclude (e.g., possibly using Remark 22.5) that the map $\mathrm{End}_k(V) \to k[G]$ is also $G \times G$-equivariant, provided we use the second action of $G \times G$ on $\mathrm{End}_k(V)$, (89) (i.e., $(g_1, g_2) \cdot A = \rho(g_2) A \rho(g_1)^{-1}$). But verify this directly as well.

(iv) In contrast, verify (a previous claim from this lecture) that the 'action map' $k[G] \to \bigoplus_{i=1}^{r} \mathrm{End}_k(V_i)$ is only $G \times G$-equivariant if we give the first action of of $G \times G$ on $\mathrm{End}_k(V)$ (i.e., $(g_1, g_2) \cdot A = \rho(g_1) A \rho(g_2)^{-1}$). Deduce that for 'most' G (find what the 'most' here is) the 'matrix coefficient' map (93) cannot possibly be in general an inverse to (90).

(v) Note that $\mathrm{End}_k(V) \to k[G]$ sends the identity matrix to some $f \in k[G]$ such that $f(1) = \dim V$. Use this to give a different deduction, for nonabelian groups G, that the 'matrix coefficient' map (93) is not a ring homomorphism and hence cannot possibly be in general an inverse to (90).

Example 22.8. Suppose (ρ, V) is a one-dimensional representation of G. Then we can think of ρ as a character $\chi : G \to k^\times$, such that $\rho(g) \cdot v = \chi(g)v$ for each $v \in V$. In this case, each matrix coefficient $c_{u,v}$ is a multiple of χ:

$$c_{u,v} = \langle u, v \rangle \cdot \chi.$$

Thus, the matrix coefficients play a role that elements of $\mathrm{Hom}(G, k^\times)$ played in the abelian case, and are particularly nice elements in $k[G]$. We will develop on this theme further in this lecture.

22.3. **The surjectivity of the matrix coefficient map.** The surjectivity does not need k to be algebraically closed, but of course needs that $(\mathrm{char}\, k, \#G) = 1$:

Lemma 22.9. *Assume that* $(\mathrm{char}\, k, \#G) = 1$. *Let* $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ *be the irreducible representations of* G *up to isomorphism. Then the matrix coefficient map of* (93),

$$\prod_{i=1}^{r} \mathrm{End}_k(V_i) \to k[G],$$

is surjective.

Proof. Note that if (σ, W) and (σ', W') are two representations of G, then we have an identification $(W \oplus W')^\vee \cong W^\vee \oplus (W')^\vee$, and it is easy to verify that the image of the matrix coefficient map associated to $\sigma \oplus \sigma'$ is the sum of the images of the matrix coefficient maps associated to σ and σ' (if you like matrices, then the matrix of $(\sigma \oplus \sigma')(g)$ can be chosen to be 'block' diagonal with respect to a suitable basis, and $\mathrm{tr}((\sigma \oplus \sigma')(g) \cdot A) = \mathrm{tr}(\sigma(g) \cdot A_1) + \mathrm{tr}(\sigma'(g) \cdot A_2)$, where A_1 is the top left $(\dim W \times \dim W)$-block of A and A_2 is the bottom right $(\dim W' \times \dim W')$-block of A).

Therefore, it is enough to find a single finite dimensional representation (ρ, V) of G such that the matrix coefficient map $V^\vee \otimes_k V \cong \mathrm{End}_k(V) \to k[G]$ is surjective. Let $(\rho, V = k[G])$ be the right regular representation of G on $k[G]$: $(g_1 \cdot f)(g) = f(gg_1)$. To show that $f \in k[G]$ lies in the image of the matrix coefficient map associated to ρ, take $v = f \in V = k[G]$ and $u =$ (evaluation at e) $\in V^\vee$. Then the image of $u \otimes v$ under the matrix coefficient map is:

$$g \mapsto \langle u, g \cdot v \rangle = (\rho(g)f)(e) = f(g),$$

or in other words, f. □

Now we give another proof for Lemma 22.9, because it gives some motivation for the definition of the matrix coefficient map:

Alternate proof for Lemma 22.9. Since the matrix coefficient map is equivariant under $G \times G \supset \{1\} \times G$, it is enough to show that, giving $k[G]$ the right regular action, the isotypic component of each irreducible (ρ, V) in it, call it $k[G]_\rho$, is contained in the image of the matrix coefficient map. Recall that by semisimplicity (which holds as $(\operatorname{char} k, \#G) = 1$) we have a surjective evaluation map

$$ev_\rho : \operatorname{Hom}_G(V, k[G]) \otimes_k V \twoheadrightarrow k[G]_\rho,$$

sending each $\varphi \otimes v$ to $\varphi(v)$; $\operatorname{Hom}_G(V, k[G])$ is the 'multiplicity space' for V in $k[G]$. [68]

Note that we have an identification $(k[G], \text{right regular}) = \operatorname{coInd}_{\{e\}}^G k$, so by Frobenius reciprocity for coinduction, we have an identification $\operatorname{Hom}_G(V, k[G]) \cong \operatorname{Hom}_{\{e\}}(V, k) = V^\vee$, so that ev_ρ identifies with a map (still called ev_ρ)

$$ev_\rho : V^\vee \otimes_k V \to k[G]_\rho.$$

Check that ev_ρ is just the matrix coefficient map for (V, ρ), which is surjective as ev_ρ is. This also gives us a conceptual reason for why we should expect the multiplicity space' $\operatorname{Hom}_G(V, k[G])$ for V to be $V^\vee$, motivating the "$V^\vee \otimes_k V$" in the aimed-for decomposition of $k[G]$ in its own terms without involving $\operatorname{End}_k(V)$. □

In contrast, when k is not algebraically closed, the matrix coefficient map is usually not injective. Even in the algebraically closed case, we saw in (iv) and (v) of Exercise 22.7 that the matrix coefficient map (93) is not an inverse to (87). Thus, we need to modify the latter map appropriately to get an inverse.

22.4. **The averaging map.**

Notation 22.10. If V is a representation of G (over k, as usual), and $(\#G, \operatorname{char} k) = 1$, we will denote by $Av_G : V \to V^G$ the linear map $v \mapsto (\#G)^{-1} \sum_{g \in G} g \cdot v$ (recall that $V^G \subset V$ is the subspace of elements of V fixed by G). Note that it is a projection from V onto V^G.

Recall the key property of irreducible representations over algebraically closed fields: if k is algebraically closed and (ρ, V) is irreducible, then $\operatorname{End}_G(V) = k$. Rather than k being algebraically closed, this is the only property that we will mostly need:

Definition 22.11. A representation (ρ, V) of G over k is said to be absolutely irreducible if its endomorphism algebra consists of just the scalars, i.e., $k \hookrightarrow \operatorname{End}_{k[G]}(V)$ is an isomorphism.

Exercise 22.12. Show that a representation (ρ, V) of G over k is absolutely irreducible if and only if for some, or equivalently any, algebraically closed field L containing k, "ρ remains irreducible after base-change to L", i.e., the composite

$$G \xrightarrow{\rho} GL_k(V) \overset{T \mapsto T \otimes \mathrm{id}_L}{\to} GL_L(V \otimes_k L)$$

[68] This is slightly different from the earlier ev_ρ, though: this one factors through the earlier one, which was an isomorphism $\operatorname{Hom}_G(V, k[G]) \otimes_D V \to k[G]_\rho$, where D is the division algebra $\operatorname{End}_G(V)$.

is an irreducible representation of G over L.
Hint: Tensoring with L over k preserves the endomorphism algebra, since we are looking at the solution space to some linear equations over k.

Example 22.13. (i) Clearly, an absolutely irreducible representation is irreducible, but the converse is not true (see (ii) below).
(ii) Clearly, a representation of a finite abelian group is absolutely irreducible if and only if it is one dimensional. Thus, for instance, the 'rotation' representations of $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z}$ on $\mathbb{R}^2$ are not absolutely irreducible.
(iii) On the other hand, in Example 20.21 from Lecture 20, we studied the irreducible representations of S_3 over an arbitrary field. It is immediate from the classification there that any irreducible representation of S_3 over any field k is absolutely irreducible.

While the proof of Fourier inversion can at first glance seem like unpleasant book-keeping, the key innovation involved is Schur's lemma packaged into the following simpler result.

Proposition 22.14. *(i) If V is an absolutely irreducible representation of G, then $\dim V \neq 0$ in k (i.e., $\dim V$ is not divisible by $\operatorname{char} k$).*
(ii) For two irreducible representations V, W of G, and $A \in \operatorname{Hom}_k(V, W)$, we have:

$$Av_G(A) = \begin{cases} 0, & \text{if } V \ncong W, \text{ and} \\ \frac{\operatorname{tr} A}{\dim V} \cdot \mathrm{Id}, & \text{if } V = W \text{ is an absolutely irreducible representation} \end{cases}.$$

Here, Av_G is being applied to the representation $\operatorname{Hom}_k(V, W)$ of G, and no assertion is made about the case where $V = W$ is not absolutely irreducible (for more information on that case see Proposition 22.41).

Proof. We will prove both parts of the proposition simultaneously. If $V \ncong W$, then $Av_G(A) \in \operatorname{Hom}_k(V, W)^G = \operatorname{Hom}_G(V, W) = 0$, so there is nothing to prove.

So assume that $V = W$ and that V is an absolutely irreducible representation. Then $Av_G(A) \in \operatorname{End}_k(V)^G = k \cdot \mathrm{Id}_V$, this last equality holding because V is absolutely irreducible. Hence $Av_G(A) = a \cdot \mathrm{Id}$ for some scalar $a \in k$. But since the averaging process respects trace, we get:

$$\operatorname{tr} A = \operatorname{tr}(a \cdot \mathrm{Id}) = a \cdot \dim V.$$

Choosing some A so that $\operatorname{tr} A \neq 0$, we get $\dim V \neq 0$, and now both (i) and (ii) follow for general A. □

22.5. **Fourier inversion.** Now we state and prove Fourier inversion for finite nonabelian groups:

Proposition 22.15. *Assume that each irreducible representation of G is absolutely irreducible (which is automatic if k is algebraically closed), and that $(\operatorname{char} k, \#G) = 1$. Then*

the map $\bigoplus_{i=1}^r \mathrm{End}_k(V_i) \to k[G]$*, that sends each* $A_j \in \mathrm{End}_k(V_j) \subset \bigoplus_{i=1}^r \mathrm{End}_k(V_i)$ *to*

$$g \mapsto \frac{\dim V_j}{\#G} \operatorname{tr}(\rho_j(g^{-1})A_j),$$

is a two-sided inverse to $k[G] \to \bigoplus_{i=1}^r \mathrm{End}_k(V_i)$*. This recovers that* (87) *is an isomorphism of rings* $k[G] \to \prod_{i=1}^r \mathrm{End}_k(V_i)$*, and also proves it to be an isomorphism of* $G \times G$*-representations* $k[G] \to \bigoplus_{i=1}^r \mathrm{End}_k(V_i)$*.*

Notation 22.16. By the Fourier inversion map for G, we mean the map $\bigoplus_{i=1}^r \mathrm{End}_k(V_i) \to k[G]$ described in the proposition. Explicitly, it is:

$$(A_i)_{i=1}^r \mapsto \Big(g \mapsto \sum_{i=1}^r \frac{\dim V_i}{\#G} \operatorname{tr}(\rho_i(g^{-1})A_i)\Big). \tag{94}$$

The assertion of Proposition 22.15 is that the Fourier inversion map, (94), is a two-sided inverse to (90).

Remark 22.17. Note that above map corrects for the failure of the 'matrix coefficient' map (93) to be an inverse to the Fourier transform map (90) in the following two ways:

(i) By replacing the $\operatorname{tr}(\rho(g)A)$ of the matrix coefficient map with $\operatorname{tr}(\rho(g^{-1})A)$, it fixes the problem mentioned in Exercise 22.7(iv), since $g \mapsto g^{-1}$ replaces the (g_1, g_2)-action on $k[G]$ with the (g_2, g_1)-action. Note that such an 'inverse' was also seen in the abelian case of Fourier inversion: see Exercise **??** of Lecture 20, especially (86) there.

(ii) By adding in the factor $(\dim V_j)^{-1}$, it fixes the problem mentioned in Exercise 22.7(v). Note that this factor did not show up in the abelian (algebraically closed) case, as there each $\dim V_j$ was equal to 1.

(iii) We have, as vector space maps $\bigoplus_{i=1}^r \mathrm{End}_k(V_i) \to k[G]$,

$$\text{(The Fourier inversion map)} = \iota \circ \text{(the matrix coefficient map)} \circ \mathcal{T}, \tag{95}$$

where $\iota : k[G] \to k[G]$ is composition with $g \mapsto g^{-1}$, and $\mathcal{T} : \bigoplus_i \mathrm{End}_k(V_i) \to \bigoplus_i \mathrm{End}_k(V_i)$ sends $(A_i)_i$ to $\left(\frac{\dim V_i}{\#G} \cdot A_i\right)_i$. Since ι and $\mathcal{T}$ are clearly isomorphisms (use that each $\dim V_i \neq 0$ by Proposition 22.14 and the assumption of absolute irreducibility), it follows that the Fourier inversion map (94) is injective/surjective/bijective if and only if the matrix coefficient map (93) is.

(iv) Using (iii) and the fact that the matrix coefficient map is surjective (Lemma 22.9), it follows that the Fourier inversion map is surjective as well.

Proof of Proposition 22.15. It is enough to show that the composite $\mathrm{End}_k(V_j) \to k[G] \to \mathrm{End}_k(V_{j'})$ (where the first map is as given in the proposition, and the second map is ρ'_j) is 0 if $j \neq j'$, and the identity if $j = j'$: while this much only gives a left-inverse to $\bigoplus_{i=1}^r \mathrm{End}_k(V_i) \to k[G]$, this suffices by the surjectivity of the Fourier inversion map (Remark 22.17(iv)).

This translates to proving that for each $1 \leqslant j, j' \leqslant r$, and each $A \in \mathrm{End}_k(V_j)$, we have an equality in $\mathrm{End}_k(V_j')$:

$$\frac{1}{\#G}\sum_{g\in G}\mathrm{tr}(\rho_j(g^{-1})A)\rho_{j'}(g) = \begin{cases} 0, & \text{if } j \neq j', \text{ and} \\ \frac{1}{\dim V_j}A, & \text{if } j = j' \end{cases}.$$

($(\dim V_j)^{-1}$ makes sense by Proposition 22.14, since ρ_j is absolutely irreducible). This is proved in the following lemma. □

Lemma 22.18. *Assume that* $(\mathrm{char}\, k, \#G) = 1$. *Then for irreducible representations* (ρ, V) *and* (ρ', V') *of* G, *we have:*

$$\frac{1}{\#G}\sum_{g\in G}\mathrm{tr}(\rho(g^{-1})A)\rho'(g) = \begin{cases} 0, & \textit{if } \rho \ncong \rho', \textit{ and} \\ \frac{1}{\dim V}A, & \textit{if } (\rho, V) = (\rho', V') \textit{ is absolutely irreducible.} \end{cases} \tag{96}$$

Proof. Again, in the second case, $(\dim V)^{-1}$ makes sense in k by Proposition 22.14 and absolute irreducibility. It is enough to prove the claimed equality when $A = A_{u,v}$ with $u \in V^\vee$ and $v \in V$ arbitrary, since such $A_{u,v}$ span $\mathrm{End}_k(V)$. Since $\rho(g^{-1})A_{u,v} = A_{u,\rho(g^{-1})v}$, this is equivalent to showing that for all $u \in V^\vee, v \in V$ and $v' \in V$ we have:

$$\frac{1}{\#G}\sum_{g\in G}\langle u, \rho(g^{-1})v\rangle\rho'(g)v' = \begin{cases} 0, & \text{if } \rho \ncong \rho', \text{ and} \\ \frac{1}{\dim V}\langle u, v'\rangle v, & \text{if } (\rho, V) = (\rho', V') \text{ is absolutely irreducible.} \end{cases}$$

For fixed u, v' and varying $v \in V$, view both sides as a map $V \to V'$. The left-hand side is $Av_G(A_{u,v'})$, while the right-hand side is 0 or $(\dim V)^{-1}\langle u, v'\rangle \cdot \mathrm{Id} = (\dim V)^{-1}\,\mathrm{tr}(A_{u,v'}) \cdot \mathrm{Id}$ (use Remark 22.4), so we are done by Proposition 22.14(ii). □

Corollary 22.19. *The matrix coefficient map* (93) *defines an isomorphism of* $G \times G$*-representations*

$$\bigoplus_{i=1}^{r}\mathrm{End}_k(V_i) \cong \bigoplus_{i=1}^{r} V_i^\vee \otimes_k V_i \to k[G],$$

where $\mathrm{End}_k(V_i)$ *is given the second action of* $G \times G$, (89).

Proof. We are given a $G \times G$-map from the left-hand side to the right-hand side, and it is enough to prove that it is a vector space isomorphism. Since we know the analogous result for the Fourier inversion map (Proposition 22.15), the same follows for the matrix coefficient map, by Remark 22.17(iii). □

As we said, the above proof doesn't use Burnside's theorem, so we get an alternate proof for it.

Corollary 22.20. *(i) If* $\rho : G \to GL_k(V)$ *is an absolutely irreducible representation of* G, *then* $\rho : k[G] \to \mathrm{End}_k(V)$ *is surjective.*

(ii) If (ρ_1, V_1) *and* (ρ_2, V_2) *are irreducible representations of* G_1 *and* G_2, *then* $(\rho_1 \otimes_k \rho_2, V_1 \otimes_k V_2)$ *is irreducible.*

Proof. The map $k[G] \to \mathrm{End}_k(V)$ is surjective, since Proposition 22.15 gives an explicit section to it. We saw in Lecture 21 that the second assertion follows from the first (see the proof of Proposition 21.14). Note that this deduction does use ring theory, but not the results of Artin-Wedderburn, Jacobson or Burnside. □

Corollary 22.21. *Assume that each irreducible representation of G is absolutely irreducible (which is automatic if k is algebraically closed), and that $(\mathrm{char}\, k, \#G) = 1$. Then either of the Fourier inversion map or the matrix coefficient map gives a decomposition of the regular representation of $G \times G$ into irreducible representations, as:*

$$k[G] \cong \bigoplus_{i=1}^{r} \mathrm{End}_k(V_i) \cong \bigoplus_{i=1}^{r} V_i^\vee \otimes_k V_i. \tag{97}$$

In particular, each of these irreducible components appears with multiplicity 1*.*

Moreover, the left regular representation of $k[G]$ and the right-regular representation of $k[G]$ each have the following decomposition in terms of irreducible representations:

$$k[G] = \bigoplus_{i=1}^{r} V_i^{\dim V_i}.$$

We also recover the formula $\#G = \sum_{i=1}^{r} (\dim V_i)^2$.

Proof. By Proposition 22.15 and Corollary 22.19, all we need to show is that each $\mathrm{End}_k(V_i) \cong V_i^\vee \otimes_k V_i$ is irreducible as a representation of $G \times G$, which follows from Corollary 22.20. □

22.6. **Digesting Fourier inversion a bit more.** The Fourier transform is an isomorphism of rings, and (appropriately interpreted) of $G \times G$-representations. We will now show the algebraic Plancherel formula for finite nonabelian groups, that the Fourier transform respects certain natural bilinear forms on either side. When $k = \mathbb{C}$, this bilinear form will turn out to have a sesquilinear variant, an inner product, more reminiscent of the usual Plancherel formula.

Notation 22.22. (i) For $f_1, f_2 \in k[G]$, define

$$\langle f_1, f_2 \rangle = (\#G) \cdot \sum_{g \in G} f_1(g^{-1}) f_2(g).$$

Note that $\langle \cdot, \cdot \rangle$ is a symmetric nondegenerate bilinear form on $k[G]$.

(ii) If (ρ, V) is any absolutely irreducible representation of G, define $(A, B) = (\dim V) \cdot \mathrm{tr}(AB)$. Thus, under the assumption that each (ρ_i, V_i) is absolutely irreducible, this also defines a symmetric nondegenerate bilinear form on $\bigoplus_{i=1}^{r} \mathrm{End}_k(V_i)$:

$$((A_i)_i, (B_i)_i) = \sum_{i=1}^{r} (\dim V_i) \cdot \mathrm{tr}(A_i B_i).$$

We write down some formulas that the proof of Proposition 22.15 gives, for $(\rho, V), (\rho', V')$ irreducible representations of G with V absolutely irreducible:

Corollary 22.23. *Assume that* $(\operatorname{char} k, \#G) = 1$. *Let* (ρ, V) *and* (ρ', V') *be irreducible representations of* G *over* k.

(i) For all $A, B \in \operatorname{End}_k(V)$, *we have:*

$$\frac{1}{\#G}\sum_{g\in G}\operatorname{tr}(\rho(g)^{-1}A)\operatorname{tr}(\rho'(g)B) = \begin{cases} 0, & \text{if } \rho \not\cong \rho', \text{ and} \\ (\dim V)^{-1}\operatorname{tr}(AB), & \text{if } (\rho, V) = (\rho', V') \text{ is absolutely irreducible.} \end{cases}$$

(ii) (Schur orthogonality for matrix coefficients) For all $u \in V^\vee, v \in V, u' \in V'^\vee$ *and* $v' \in V'^\vee$,

$$\frac{1}{\#G}\sum_{g\in G} c_{u,v}(g^{-1})c_{u',v'}(g) = \begin{cases} 0, & \text{if } \rho \not\cong \rho', \text{ and} \\ (\dim V)^{-1}\cdot\langle u', v\rangle\langle u, v'\rangle, & \text{if } (\rho, V) = (\rho', V') \text{ is absolutely irreducible.} \end{cases}$$

(iii) If $(\rho_1, V_1), \dots, (\rho_r, V_r)$ *are all absolutely irreducible, the Fourier inversion isomorphism,* $\bigoplus_{i=1}^r \operatorname{End}_k(V_i) \to k[G]$,

$$(A_i)_i \mapsto \Big(g \mapsto \sum_{i=1}^r \frac{\dim V_i}{\#G}\operatorname{tr}(\rho_i(g^{-1})A_i)\Big)$$

transports the form $(\cdot,\cdot)$ *on* $\bigoplus_{i=1}^r \operatorname{End}_k(V_i)$ *to the form* $\langle\cdot,\cdot\rangle$ *on* $k[G]$.

(iv) (Plancherel formula) Equivalently (looking in the opposite direction), if $(\rho_1, V_1), \dots, (\rho_r, V_r)$ *are absolutely irreducible, the action map* $k[G] \to \bigoplus_{i=1}^r \operatorname{End}_k(V_i)$ *transports the form* $\langle\cdot,\cdot\rangle$ *on* $k[G]$ *to the form* $(\cdot,\cdot)$ *on* $\bigoplus_{i=1}^r \operatorname{End}_k(V_i)$*: for all* $f_1, f_2 \in k[G]$ *we have:*

$$(\#G)\cdot\sum_{g\in G} f_1(g^{-1})f_2(g) = \sum_{i=1}^r (\dim V_i)\cdot\operatorname{tr}(\rho_i(f_1)\rho_i(f_2)).$$

(Note also that $\rho_i(f_1)\rho_i(f_2) = \rho_i(f_1 * f_2)$*).*

Proof. For (i), multiply the formula of Lemma 22.18 (see (96)) by B, and apply tr.

For (ii), since $c_{u,v}(g) = \operatorname{tr}(\rho(g)A_{u,v})$ and similarly with $c_{u',v'}$, we can simply apply (i) with $A = A_{u,v}$ and $B = A_{u',v'}$, and note that when $(\rho, V) = (\rho', V')$,

$$\operatorname{tr}(AB) = \operatorname{tr}(A_{u,v}A_{u',v'}) = \langle u', v\rangle\operatorname{tr}(A_{u,v'}) = \langle u', v\rangle\langle u, v'\rangle,$$

where we used the equality $A_{u,v}\circ A_{u',v'} = \langle u', v\rangle A_{u,v'}$, which is justified as follows:

$$A_{u,v}(A_{u',v'}(w)) = \langle u, A_{u',v'}w\rangle v = \langle u, \langle u', w\rangle v'\rangle v = \langle u', w\rangle\langle u, v'\rangle v = \langle u, v'\rangle A_{u',v}(w).$$

For (iii), note that $\langle\cdot,\cdot\rangle$ applied to the images of $(A_i)_i$ and $(B_i)_i$ equals:

$$(\#G)\sum_{g\in G}\Big(\sum_{i=1}^r \frac{\dim V_i}{\#G}\operatorname{tr}(\rho_i(g)A_i)\Big)\Big(\sum_{i=1}^r \frac{\dim V_i}{\#G}\operatorname{tr}(\rho_i(g^{-1})B_i)\Big),$$

which by the first case of (i) equals

$$(\#G)\sum_{g\in G}\Big(\sum_{i=1}^{r}\frac{(\dim V_i)^2}{(\#G)^2}\operatorname{tr}(\rho_i(g^{-1})B_i)\operatorname{tr}(\rho_i(g)A_i)\Big),$$

which in turn, by the second case of (i), equals

$$\sum_{i=1}^{r}(\dim V_i)\operatorname{tr}(B_iA_i)=\sum_{i=1}^{r}(\dim V_i)\operatorname{tr}(A_iB_i)=((A_i)_i,(B_i)_i).$$

This gives (iii).

Finally, (iv) is an immediate consequence of (iii): if an isomorphism of vector spaces transports a bilinear form B_1 to a bilinear form B_2, its inverse transports B_2 to B_1. □

Remark 22.24. Schur orthogonality for matrix coefficients, namely Corollary 22.23(ii), gives us a nice basis of functions for $k[G]$ that behaves well with respect to the form $\langle\cdot,\cdot\rangle$: for $1\leqslant i\leqslant r$, if $e_{i,1},\dots,e_{i,n_i}$ is a basis of V_i and $e_{i,1}^\vee,\dots,e_{i,n_i}^\vee$ is the dual basis of $V_i^\vee$, then $(c_{i,p,q}:=c_{e_{i,p},e_{i,q}})_{1\leqslant i\leqslant r,1\leqslant p,q\leqslant n_i}$ is a basis for $k[G]$, which by Corollary 22.23(ii) satisfies:

$$\big\langle c_{i,p,q},c_{i',p',q'}\big\rangle=\begin{cases}0, & \text{if } i\neq i', \text{ or if } i=i' \text{ and } (p,q)\neq(q',p'), \text{ and}\\ \frac{(\#G)^2}{\dim V_i}, & \text{otherwise}\end{cases}.$$

Thus, these don't form an orthogonal basis, but almost, in that $c_{i,p,q}$ pairs only with $c_{i,q,p}$. Recall that when G was abelian and k was algebraically closed (with $(\operatorname{char}k,\#G)=1$), we got an extremely nice basis for $k[G]$, namely $\operatorname{Hom}(G,k^\times)\subset k[G]$. The above $c_{i,p,q}$ is pretty much the best substitute we can have in the nonabelian case. Note that for a fixed i, $\operatorname{Span}(\{c_{i,p,q}\mid 1\leqslant p,q\leqslant n_i\})\subset k[G]$ is an irreducible $(G\times G)$-subrepresentation, and also the $\rho_i^\vee\otimes\rho_i$-isotypic component. It is also the $\rho_i^\vee$-isotypic component for the left regular representation, which is the ρ_i-isotypic component for the right regular representation.

22.7. **Inner product versions over the complex numbers.** This is optional, but I recommend that you at least quickly glance through it.

Proposition 22.25. *Let $k=\mathbb{C}$. Give G measure dg equal to $(\#G)^{-1}$ times the counting measure, and consider $L^2(G)$.*

- *(i) (Orthonormal basis for $\mathbb{C}[G]$) For each of the irreducible representations $(\rho_1,V_1),\dots,(\rho_r,V_r)$ of G, choose a G-invariant inner product on V by averaging, and let $v_{i,1},\dots,v_{i,n_i}$ be an orthonormal basis of V_i for this inner product. For $1\leqslant i\leqslant r$ and $1\leqslant p,q\leqslant n_i$, set $c_{i,p,q}(g)=\sqrt{\dim V}\langle v_{i,p},\rho(g)v_{i,q}\rangle$. Then $\{c_{i,p,q}\mid 1\leqslant i\leqslant r,1\leqslant p,q\leqslant n_i\}$ is an orthonormal basis for $L^2(G,dg)$.*
- *(ii) (Plancherel formula) Make each $\operatorname{End}_{\mathbb{C}}(V_i)$ into a Hilbert space by giving it $\dim V_i$ times the Hilbert-Schmidt norm $\|\cdot\|_{HS}$, where $\|A_i\|_{HS}=\operatorname{tr}(A_i^*A_i)=\sum_{1\leqslant p,q\leqslant n_i}\langle A_ie_{i,p},A_ie_{i,q}\rangle$*

for any orthonormal basis $\{e_{i,1}, \dots, e_{i,n_i}\}$ of V_i. Then the normalized Fourier transform map

$$L^2(G) = \mathbb{C}[G] \to \bigoplus_{i=1}^{r} \mathrm{End}_{\mathbb{C}}(V_i), \qquad \Big(f \mapsto ((\#G)^{-1}\rho_i(f))_{i=1}^r = \Big(\frac{1}{\#G}\sum_{g\in G} f(g)\rho_i(g)\Big)_{i=1}^r\Big)$$

(i.e., we multiplied the usual Fourier transform by $(\#G)^{-1}$ to account for our change of measure on G), is a Hilbert space isomorphism.

Proof. First we prove (i). Note that $\langle v_{i,p}, - \rangle \in V_i^\vee$ for each i and p, so each $c_{i,p,q}$ is a matrix coefficient for (ρ_i, V_i). Now we have, by Corollary 22.23(iii),

$$\frac{1}{\#G}\sum_{g\in G}(\sqrt{\dim V_i}c_{i,p,q}(g^{-1}))(\sqrt{\dim V_j}c_{i,p',q'}(g)) = \begin{cases} 0, & \text{if } i \neq j, \text{ and} \\ \langle v_{i,p}, v_{i,q'}\rangle\langle v_{i,p'}, v_{i,q}\rangle, & \text{otherwise.} \end{cases}.$$

Now (i) follows on noting that, since the action of g^{-1} on V_i is unitary, we have

$$c_{i,p,q}(g^{-1}) = \sqrt{\dim V_i}\langle v_{i,p}, g^{-1}\cdot v_{i,q}\rangle = \sqrt{\dim V_i}\langle g\cdot v_{i,p}, v_{i,q}\rangle = \overline{\sqrt{\dim V_i}\langle v_{i,q}, g\cdot v_{i,p}\rangle} = \overline{c_{i,q,p}(g)}.$$

For (ii), we apply Corollary 22.23(iv), replacing $f_1(g)$ with $\overline{f_1(g^{-1})}$. Then $(\#G)\rho(f_1) = \sum_{g\in G} f_1(g)\rho(g)$ gets replaced with

$$\sum_{g\in G}\overline{f_1(g^{-1})}\rho(g) = \sum_{g\in G}\overline{f_1(g)}\rho(g^{-1}) = \sum_{g\in G}\overline{f_1(g)}\rho(g)^* = (\#G)(\rho_1(f))^*,$$

and then the equality given by Corollary 22.23(iv) becomes simply the one given by (ii) (up to a factor of $(\#G)^2$ on both sides). □

Exercise 22.26. Work out how the constructs defined in Proposition 22.25 change when we change the choice of the inner products on the V_i (exercise: they are well-defined up to positive scalars), and explicate how the validity of the proposition remains unaffected by these changes.

Remark 22.27. Proposition 22.25 generalizes to compact (Hausdorff) topological groups, with some obvious modifications. *The following will skip most details and likely mess up the rest, especially the constants, but for a mistake-free version with more details you can see https://terrytao.wordpress.com/2011/01/23/the-peter-weyl-theorem-and-non-abelian-fourier-analysis-on-compact-groups/ .*

One gets, with μ_G the normalized Haar measure on G:

$$L^2(G, \mu_G) \cong \hat{\bigoplus}_{V\in Irr(G)}(\mathrm{End}_k(V), (\dim V)\cdot\|\cdot\|_{HS}) = \hat{\bigoplus}_{V\in Irr(G)}(V^\vee \otimes V, (\dim V)\cdot\|\cdot\|),$$

where one identifies $V^\vee \otimes V$ with "$\bar{V}\otimes_{\mathbb{C}} V$", $\bar{V} = \mathbb{C}\otimes_{\mathbb{C}} V$ where $\mathbb{C}$ is viewed as a $\mathbb{C}$-algebra via complex conjugation, and uses on $V^\vee \otimes_{\mathbb{C}} V$ the product of $\dim V$ and an 'obvious' inner product. Moreover, $Irr(G)$ denotes the set of *continuous* (topologically) irreducible representations of G on Hilbert spaces up to isomorphism, though they can all be shown to be finite dimensional and hence abstractly irreducible. Note the use of the "completed

Hilbert space direct sum" $\hat{\bigoplus}$, since the direct sum of infinitely many Hilbert spaces is not a Hilbert space, and only becomes one on completion (this matters since $Irr(G)$ is typically infinite even when G is compact). In contrast, we don't need to put a $V^\vee \hat{\otimes}_{\mathbb{C}} V$, since each V that occurs above can be shown to be finite dimensional.

In the proof of this result, the analogue of the proof of Proposition 22.15 works out pretty much analogously, but more care is needed for the analogue of Lemma 22.9. The proof of Lemma 22.9 does adapt to show that if a function $f : G \to \mathbb{C}$ belongs to a finite subrepresentation of the right-regular representation, then f is a span of matrix coefficients, but the problem is to show that such f are actually dense in $L^2(G)$. The idea is then to use the *left regular* action of some $\varphi : G \to \mathbb{C}$ to construct a compact self-adjoint operator on $L^2(G)$ that commutes with the right regular action, and use the fact that such an operator has finite dimensional nonzero eigenspaces (the spectral theorem for compact self-adjoint operators), giving a good supply of finite dimensional subspaces of the right regular representation. This could be considered one form of the Peter-Weyl theorem.

One can also show that the span of the matrix coefficients (each of which can be shown to be continuous – even smooth if G is a Lie group – without much difficulty) is dense in the Banach space $C(G)$ of continuous functions on G with respect to the supremum norm: this seems to be what is more commonly referred to as the Peter-Weyl theorem.

Remark 22.28. In Remark 22.24, I did not do a good job of justifying that the matrix coefficients $c_{i,p,q}$ are a very nice basis for $k[G]$, other than that in the abelian case they are multiplicative functions on G. But the complex case can give some hints as to why. I will give some vague explanations in this remark. While I haven't read/worked out the details, I would like to quote and partially make the case that, when G is a compact Lie group, the matrix coefficients are solutions of suitable differential equations. As I said above, the matrix coefficients are smooth, so differential operators can be applied to them. The point is that while G itself may not have much of a center, there are always numerous differential operators that commute with the action of G, and hence by Schur's lemma act by scalar multiplication on irreducible representations and hence on matrix coefficients! Here, one uses that due to finite dimensionality, irreducible representations consist of 'smooth vectors' that can be differentiated.

As examples, look at the Fourier series and the Fourier transform. The $x \mapsto e^{inx}$ and the $x \mapsto e^{iyx}$ are eigenfunctions for d/dx, with eigenvalues in and iy, respectively. This is precisely because in each case the differential operator id/dx or d/dx was an 'infinitesimal version' of the group action, and commuted with the action of G because (in this case) G itself was abelian. Thus, in a sense, Schur's lemma is sort of responsible for why the classical Fourier transform involves eigenfunctions for d/dx. Apparently many 'special functions' can be explained this way. Another example consists of 'spherical harmonics', which are matrix coefficients for the compact group SO_3 associated to its left regular action on $L^2(S^2) = L^2(SO_3/SO_2)$.

22.8. **Schur orthogonality relations for characters and for conjugacy classes.**

Definition 22.29. Henceforth, if (ρ, V) is a (finite dimensional) representation of G, $\chi_\rho = \chi_V \in k[G]$ will denote the map $G \to k$ such that for all $g \in G$,

$$\chi_\rho(g) = \chi_V(g) = \operatorname{tr} \rho(g).$$

χ_ρ will be called the character of (ρ, V). When we write $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ for the irreducible representations of G up to isomorphism, we may write χ_i for χ_{ρ_i}.

Corollary 22.30 (Schur orthogonality for characters)**.** *(i) If V, W are irreducible representations of k, then*

$$\langle \chi_V, \chi_W \rangle = \begin{cases} 0, & \text{if } V \not\cong W, \text{ and} \\ 1, & \text{if } V = W \text{ is an absolutely irreducible representation} \end{cases}.$$

(ii) Suppose that $k = \bar{k}$. Then for any two representations V, W of G, $\langle \chi_V, \chi_W \rangle = \dim \operatorname{Hom}_G(V, W)$ in k (note that this may involve loss of information when $\operatorname{char} k \neq 0$).

Proof. The first assertion follows from Corollary 22.23(i) by taking $A = \mathrm{Id}_V$ and $B = \mathrm{Id}_W$.

The second assertion follows from the first: use that both sides are additive by semisimplicity, and that irreducible is the same as absolutely irreducible when k is algebraically closed.

However, these special cases of Fourier inversion are also easy to handle without going through all that went into Corollary 22.23 (and hence in many sources they directly go to characters without involving the matrix coefficients), so let us see them directly. Show as an easy exercise that $\chi_{V^\vee}(g) = \chi_V(g^{-1})$ for each $g \in G$. We get:

$$\frac{1}{\#G} \sum_{g \in G} \chi_V(g^{-1}) \chi_W(g) = \frac{1}{\#G} \sum_{g \in G} \chi_{V^\vee}(g) \chi_W(g) = \frac{1}{\#G} \sum_{g \in G} \chi_{V^\vee \otimes W}(g)$$
$$= \frac{1}{\#G} \sum_{g \in G} \chi_{\operatorname{Hom}(V,W)}(g) = \operatorname{tr}(Av_G|_{\operatorname{Hom}_k(V,W)}) = \dim_G \operatorname{Hom}_G(V, W),$$

since Av_G on $\operatorname{Hom}_k(V, W)$ is a projection onto $\operatorname{Hom}_G(V, W) \subset \operatorname{Hom}_k(V, W)$.

The above computation used the implicit assumption that either V is absolutely irreducible or k is algebraically closed, depending on which assertion one wants to prove. □

Notation 22.31. Given $g \in G$, $Z_G(g)$ will denote the centralizer of g and $C(g)$ its conjugacy class.

Theorem 22.32 (Schur orthogonality relations for conjugacy classes)**.** *Assume that each irreducible representation of G is absolutely irreducible. If $g, h \in G$, then*

$$\sum_{\chi \in Irr(G)} \chi(g^{-1}) \chi(h) = \begin{cases} 0, & \text{if } g, h \text{ are not } G\text{-conjugate, and} \\ \#Z_G(g), & \text{otherwise.} \end{cases}$$

The proof of this theorem will use the following observation, which will be useful elsewhere as well:

Lemma 22.33. *Let $f = \mathbb{1}_C \subset Z(k[G])$, where $C \subset G$ is a conjugacy class. Then for any absolutely irreducible representation (ρ, V) of G, $\rho(f) \in \mathrm{End}_k(V)$ is multiplication by the scalar*

$$\lambda = \frac{\#C_g}{\dim V} \cdot \chi_\rho(g).$$

Proof. Since f is conjugation invariant, $\rho(f) \circ \rho(g) = \rho(g) \circ \rho(f)$ for all $g \in G$, and hence we indeed have $f \in Z(k[G])$. By Schur's lemma, f acts on V by a scalar λ (this uses absolute irreducibility). If $\chi = \chi_\rho$, by taking traces, we get:

$$(\dim_k V) \cdot \lambda = \sum_{h \in G} \mathbb{1}_{C_g}(h) \operatorname{tr} \rho(h) = \sum_{h \in C_g} \chi_\rho(h) = \#C_g \cdot \chi_\rho(g),$$

using the conjugation invariance of trace, from which the lemma follows. □

Proof of Theorem 22.32. The idea is to use Corollary 22.23(iv), taking f_1 to be the characteristic function $\mathbb{1}_{C_{g^{-1}}}$ of the conjugacy class of g^{-1}, and f_2 to be the characteristic function $\mathbb{1}_{C_h}$ of the conjugacy class of h.

By Lemma 22.33, for each i, $\rho(f_1)$ and $\rho(f_2)$ are scalar multiplications on V_i respectively by $\#C(g^{-1})(\dim V_i)^{-1}\chi_i(g^{-1}) = \#C(g)(\dim V_i)^{-1}\chi_i(g^{-1})$ and $\#C(h)(\dim V_i)\chi_i(h)$. Hence the right-hand side of the equality given by Corollary 22.23(iv) is:

$$\sum_{i=1}^{r} (\dim V_i) \operatorname{tr}(\rho_i(f_1)\rho_i(f_2)) = \#C(g)\cdot\#C(h)\cdot\sum_{i=1}^{r} \operatorname{tr}\Big(\frac{1}{\dim V_i}\chi_i(g^{-1})\chi_i(h)\cdot \mathrm{Id}\Big) = \#C(g)\cdot\#C(h)\cdot\sum_{i=1}^{r} \chi_i(g^{-1})\chi_i(h).$$

Thus, we conclude that

$$\#C(g)\cdot\#C(h)\sum_{i=1}^{r} \chi_i(g^{-1})\chi_i(h) = (\#G)\Big(\sum_{g \in G} f_1(g^{-1})f_2(g) = \begin{cases} 0, & \text{if } C_g \neq C_h, \text{ and} \\ (\#G)\cdot(\#C(g)), & \text{otherwise.} \end{cases}.$$

From this, the lemma follows, since $\#Z_G(g) = (\#G)/(\#C(g))$.

□

Recall from Lecture 21 that $Z(k[G]) \subset k[G]$ is the subspace – indeed, subalgebra – of class functions. Under the isomorphism $k[G] \to \prod_{i=1}^r \mathrm{End}_{D_i}(V_i)$, it maps to $\prod_{i=1}^r K_i$, where K_i is the center of D_i.

Proposition 22.34. *(i) If either* $\operatorname{char} k = 0$ *or k is algebraically closed, $\chi_{V_1}, \ldots, \chi_{V_r}$ are linearly independent.*

(ii) Assume that $k = \bar{k}$ is algebraically closed. Then $\chi_{V_1}, \ldots, \chi_{V_r}$ is a basis for $Z(k[G])$.

(iii) Assume that $k = \bar{k}$ is algebraically closed. Then the number of irreducible characters equals the number of conjugacy classes.

Proof. In the case where k is algebraically closed, it is easy to see (i) from Corollary 22.30. If $\operatorname{char} k = 0$, then base-changing to an algebraic closure $\bar{k}$ of k (we can base-change to compute the trace), Corollary 22.30(ii) gives:

$$\frac{1}{\#G}\sum_{g\in G}\chi_\rho(g^{-1})\chi_\rho(g') = \begin{cases} 0, & \text{if } \rho \ncong \rho', \text{ and} \\ \dim_k \operatorname{Hom}_G(V_{\bar{k}}, V_{\bar{k}}) \neq 0, & \text{if } (\rho, V) = (\rho', V'). \end{cases}$$

From this, (i) is easy to see for this case (characteristic zero) as well.

Now assume that k is algebraically closed. Though (ii) follows from (i), let us prove it directly. Taking $A = \mathrm{Id}_{V_i}$ in Lemma 22.18, we see that under the action map, $g \mapsto (\#G)^{-1}\chi_i(g^{-1})$ annihilates all the V_j with $j \neq i$, and acts as $(\dim_k V_i)^{-1}$ on V_i. Thus, using that the action map is an isomorphism $k[G] \to \prod_{i=1}^r \operatorname{End}_k(V_i)$, which restricts to an isomorphism from the centre $Z(k[G]) \subset k[G]$ consisting of the class functions to $\prod_{i=1}^r k \subset \prod_{i=1}^r \operatorname{End}_k(V_i)$, it follows that the $g \mapsto (\#G)^{-1}\chi_i(g^{-1})$ form a basis for the class functions, and hence so do the $g \mapsto \chi_i(g)$.

Finally, (iii) follows from the observation, proved in Lecture 21, that $Z(k[G])$ has a basis consisting of the characteristic functions of the conjugacy classes. □

22.9. **Dimensions of irreducible representations divide the order of the group.**

Notation 22.35. Henceforth, we will also write $Irr(G)$ in place of $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ for a set of representatives for the isomorphism classes of irreducible representations of G.

Definition 22.36. Let k be a field of characteristic zero. An element $\alpha \in k$ is said to be an algebraic integer if it satisfies the following equivalent conditions:

(i) α satisfies a monic polynomial $x^n + \sum_{i=1}^n a_i x^{n-i}$, where $a_i \in \mathbb{Z}$ for all i.
(ii) $\mathbb{Z}[\alpha] \subset k$ is a finitely generated module over $\mathbb{Z}$.
(iii) $\mathbb{Z}[\alpha] \subset k$ is contained in a finitely generated module over $\mathbb{Z}$.

The equivalence of the three conditions is easy, and is given in Lemma 22.37 below.

Lemma 22.37. *The three conditions in Definition 22.36 are equivalent.*

Proof. If (i) is satisfied, then $\{1, \alpha, \ldots, \alpha^{n-1}\}$ $\mathbb{Z}$-spans $\mathbb{Z}[\alpha]$, and (ii) follows. If (ii) is satisfied, then any given finite set of generators for $\mathbb{Z}[\alpha]$ is contained in the span of some $\{1, \alpha, \ldots, \alpha^{n-1}\}$, so that α^n can be written as a $\mathbb{Z}$-linear combination of $1, \alpha, \ldots, \alpha^{n-1}$, giving (i). Thus, (i) and (ii) are equivalent.

(ii) trivially implies (iii), and the converse follows from the fact that $\mathbb{Z}$ is Noetherian. □

Now assume $\operatorname{char} k = 0$. We have a natural basis for $Z(k[G])$, namely, the $\mathbb{1}_C$ as C varies over the conjugacy class of G.

Lemma 22.38. *$Z(k[G])_0 := Span_{\mathbb{Z}}(\{\mathbb{1}_C\}_C)$ is closed under multiplication (i.e., convolution), and is hence a subring of $Z(k[G])$ which is finitely generated as a $\mathbb{Z}$-module.*

Proof. On choosing $g(C'') \in C''$ for each conjugacy class $C'' \subset G$, this follows from the easily verified formula:

$$\mathbb{1}_C \cdot \mathbb{1}_{C'} = \sum_{C''} (\#\{g \in C, g' \in C' \mid gg' = g(C'')\}) \cdot \mathbb{1}_{C''}$$

(because this expression is independent of the choice of the $g(C'')$). □

Corollary 22.39. *If $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{C_i}$, where each $\alpha_i \in k$ is an algebraic integer and each $C_i \subset G$ is a conjugacy class, then f acts on each absolutely irreducible representation (ρ, V) of G by an algebraic integer.*

Proof. For $1 \leqslant i \leqslant n$, since α_i satisfies a monic polynomial with coefficients in $\mathbb{Z} \subset \mathbb{Z}[\alpha_1, \ldots, \alpha_{i-1}]$, we conclude that $\mathbb{Z}[\alpha_1, \ldots, \alpha_i]$ is a finitely generated module over $\mathbb{Z}[\alpha_1, \ldots, \alpha_{i-1}]$. Applying this inductively, each $\mathbb{Z}[\alpha_1, \ldots, \alpha_i]$ is a finitely generated module over $\mathbb{Z}$, and in particular so is $A := \mathbb{Z}[\alpha_1, \ldots, \alpha_n]$. It follows from Lemma 22.38 that $Z(k[G])_{0,A} := \mathrm{Span}_A(\{\mathbb{1}_C\}_C)$ is a subring of $Z(k[G])$, and clearly it is a finitely generated $\mathbb{Z}$-module.

If (ρ, V) is an absolutely irreducible representation of G, then each element $f' \in Z(k[G])_{0,A}$ acts on V as a scalar $\lambda(f')$. $f' \mapsto \lambda(f')$ is a homomorphism $\lambda : Z(k[G])_{0,A} \to k$, whose image is a subring of k which is a finitely generated $\mathbb{Z}$-module, and hence consists of algebraic integers. Thus, each $f' \in Z(k[G]_{0,A})$ acts on V by an algebraic integer $\lambda(f')$, and in particular the same applies to f. □

Proposition 22.40 (Frobenius)**.** *Suppose $k = \bar{k}$ is algebraically closed, and k has characteristic zero. Then for each irreducible representation (ρ, V) of G, $\dim V$ divides $\#G$.*

Proof. The trick is to construct an element of $R = \sum_i \alpha_i \mathbb{1}_{C_i} \in k[G]$, where each α_i is an algebraic integer, that acts on V_i by $(\#G)(\dim V_i)^{-1}$ for each $1 \leqslant i \leqslant r$.

Let $C_1, \ldots, C_r$ be the conjugacy classes of G (there are r of them, by Proposition 22.34(iii)), and let $g_i \in C_i$ for each i. Set $R = \sum_{j=1}^{r} \chi_{V_i}(g_j^{-1}) \mathbb{1}_{C_j}$. Then by Lemma 22.33, R acts on V_i by the scalar

$$\frac{1}{\dim V_i} \sum_{j=1}^{r} \#C_j \cdot \chi_{V_i}(g_j^{-1}) \chi_{V_i}(g_j) = \frac{1}{\dim V_i} \sum_{g \in G} \chi_{V_i}(g^{-1}) \chi_{V_i}(g) = \frac{\#G}{\dim V_i},$$

where we also used the Schur orthogonality for characters. It follows from Lemma 22.39 that $(\#G)/(\dim V_i)$ is an algebraic integer, and it is easy to see that any rational number which is an algebraic integer is an integer (here, we are implicitly using that $\mathbb{Q}$ embeds into k, which is true because k has characteristic zero). Therefore, $(\#G)/(\dim V_i)$ is an integer, as desired. □

22.10. **Appendix: orthogonality relations in the non-algebraically closed case.** This subsection is optional, and was not discussed in the lecture. These results (or rather their corrected versions) are certainly there somewhere in the literature, but I don't remember seeing them anywhere. So be careful believing anything here.

We continue to assume that $(\operatorname{char} k, \#G) = 1$, but do not assume that k is algebraically closed, or that the representations of interest are absolutely irreducible.

First we generalize Proposition 22.14 to the present case.

Proposition 22.41. *Let V be an irreducible representation of G over k. Let $D = \operatorname{End}_G(V)$ be the associated division algebra, and $K \supset k$ the center of D.*

(i) $\dim_K V \neq 0$.
(ii) For all $A \in \operatorname{End}_D(V)$, we have $Av_G(A) = \frac{\operatorname{tr}_K A}{\dim_K V} \cdot \mathrm{Id}$.

Proof, modulo a nontrivial fact. We will assume the following nontrivial fact from the theory of central simple algebras: D, being a central simple algebra over the bigger field K, has a maximal subfield L, with the property that $[D : K] = [L : K]^2$.

Note that V is naturally a vector space over $D = \operatorname{End}_G(V)$, and hence has compatible vector space structures over L, K and k. Thus, $\dim_K V = (\dim_L V) \cdot [L : K]$, but $\dim_L V = [L : K] \dim_D V$, so to prove (i) it suffices to show that $\dim_L V \neq 0$. Since the action of G commutes with $D \supset L$, ρ can be thought of as a representation $G \to GL_L(V) \subset GL_D(V)$.

Since the commutant of $k[G]$ in $\operatorname{End}_k(V)$ is D, the commutant of $L[G]$ in $\operatorname{End}_L(V)$ is contained in the centralizer of L in D, which is L itself (since $L \subset D$ is a maximal subfield). Thus, $G \to GL_L(V)$ is an absolutely irreducible representation of G over L, and therefore $\dim_L V \neq 0$ by Proposition 22.14(i), giving (i).

Now suppose $A \in \operatorname{End}_D(V)$. Then $Av_G(A) \in \operatorname{End}_D(V)$, since each $\rho(g)$ commutes with D. Moreover, $Av_G(A)$ commutes with each $\rho(g)$, and hence with $\rho(k[G]) = \operatorname{End}_D(V)$. Therefore, there exists $a \in K$ such that $Av_G(A) = a \cdot \mathrm{Id}$. Now take tr_K (i.e., trace of both sides as K-vector space maps):

$$\operatorname{tr}_K(A) = a \cdot \dim_K V,$$

so $a = (\dim_K V)^{-1} \operatorname{tr}_K(A)$. □

Proposition 22.42. *Let (ρ, V) and (ρ', V') be irreducible representations of G over k. Associate to V the division algebra $D = \operatorname{End}_G(V)$, and the finite extension $K = Z(D)$ of k. Thus, D is a central division algebra over K. We have for all $A \in \operatorname{End}_D(V) \subset \operatorname{End}_k(V)$:*

$$\frac{1}{\#G} \sum_{g \in G} \operatorname{tr}(\rho(g^{-1}A))\rho'(g) = \begin{cases} 0, & \text{if } \rho \ncong \rho', \text{ and} \\ \frac{1}{\dim_D V} A, & \text{if } (\rho, V) = (\rho', V'). \end{cases}$$

Proof, modulo two standard facts from field theory. If $\rho \ncong \rho'$, the proof is the same as before, so we assume that $(\rho, V) = (\rho', V')$. Consider $\operatorname{End}_D(V)$. Since each $\rho(g) \in GL_D(V)$,

$$A \mapsto \frac{1}{\#G} \sum_{g \in G} \operatorname{tr}(\rho(g^{-1}A))\rho'(g) = \frac{1}{\#G} \sum_{g \in G} \operatorname{tr}(\rho(g^{-1}A))\rho(g)$$

is an endomorphism of $\operatorname{End}_D(V)$, which is further verified to be $G \times G$-equivariant. Now we use the Jacobson density theorem (which we managed to not use in the absolutely

irreducible case), that $\rho(k[G]) = \mathrm{End}_D(V)$. Therefore, the above morphism is $\mathrm{End}_D(V) \times \mathrm{End}_D(V)$-equivariant, and since $\mathrm{End}_D(V)$ is a simple ring, it is immediately seen to be given by $A \mapsto aA$, where $a \in Z(\mathrm{End}_D(V)) = K$, i.e., it is given by multiplication by some $a \in K$ in V, which is a vector space over D and hence over K.

Let $\bar{k}$ be an algebraic closure of k. Since $K \otimes_k \bar{k} \subset Z(\mathrm{End}_D(V) \otimes_k \bar{k}) \subset \bar{k}[G] \cong \prod_{i=1}^{r'} M_{n_i}(\bar{k}_i)$ for some of r', n_i etc., it follows that that $K \otimes_k \bar{k}$ is reduced. This implies, by a general fact from field theory, that K/k is separable. Another fact that we will assume is that, in this case, the k-bilinear form $\mathrm{tr}_{K/k} : K \times K \to k$, given by $(x, y) \mapsto \mathrm{tr}_{K/k}(xy)$, is nondegenerate.

For all $b \in K$, consider the map $\mathrm{End}_D(G) \to \mathrm{End}_D(G)$, given by

$$T_b : A \mapsto \frac{1}{\#G} \sum_{g \in G} \mathrm{tr}(\rho(g^{-1})bA)\rho(g),$$

so that $T_b(A) = abA$.

Take $A = \mathrm{Id}$ to be the identity, so $\mathrm{tr}(T_b(A)) = \mathrm{tr}_k(v \mapsto abv) = (\dim_K V) \cdot \mathrm{tr}_{K/k}(ab)$. Then

$$(\dim_K V) \cdot \mathrm{tr}_{K/k}(ab) = \mathrm{tr}(T_b(\mathrm{Id})) = \frac{1}{\#G} \sum_{g \in G} \mathrm{tr}(\rho(g^{-1}b)) \, \mathrm{tr}(\rho(g)) = \frac{1}{\#G} \, \mathrm{tr}_{K/k}\Big(\mathrm{tr}_K(\rho(g^{-1}b)) \, \mathrm{tr}_K(\rho(g)) \Big).$$

Hence

$$(\dim_K V) \cdot \mathrm{tr}_{K/k}(ab) = \mathrm{tr}_{K/k}\Big(b \cdot \frac{1}{\#G} \sum_{g \in G} \mathrm{tr}_K(\rho^\vee(g) \otimes \rho(g)) \Big)$$

$$= \mathrm{tr}_{K/k}\Big(b \cdot \mathrm{tr}_K\Big(\frac{1}{\#G} \sum_{g \in G} \Big(A \mapsto \rho(g) A \rho(g^{-1}) \Big) |_{\mathrm{End}_K(V)} \Big) \Big) = \mathrm{tr}_{K/k}(b \cdot \mathrm{tr}_K(Av_G|_{\mathrm{End}_K(V)})).$$

On $\mathrm{End}_K(V)$, Av_G is a projection to $\mathrm{End}_{K[G]}(V) \cong D$, so we get

$$(\dim_K V) \cdot \mathrm{tr}_{K/k}(ab) = \mathrm{tr}_{K/k}(b \cdot \mathrm{tr}_K(Av_G(\mathrm{End}_K(V)))) = \mathrm{tr}_{K/k}(b \dim_K D) = (\dim_K D) \cdot \mathrm{tr}_{K/k}(b).$$

Therefore, using that $\dim_K V \neq 0$ in K (Proposition 22.41), and that $\mathrm{tr}_{K/k}$ is a nondegenerate bilinear form $K \times K \to k$, we get that $a = (\dim_K V)^{-1} \dim_K D = (\dim_D V)^{-1}$. $\square$

Now we give a strategy for a possible alternate proof that (hopefully) works at least in characteristic zero; I haven't tried to simplify the exposition, so it is likely extra painful to read, but this proof follows an approach that tells us how representations reduce when we base-change to an algebraic closure.

Strategy for an alternate proof for Proposition 22.42, $\mathrm{char}\, k = 0$ *for simplicity.* It is given that $\mathrm{End}_G(V) = D$. Base-changing to an algebraic closure $\bar{k}$ of k, we get a (usually reducible) representation $(\rho \otimes_k \bar{k}, V \otimes_k \bar{k})$. To see how this representation reduces, let $D \otimes_k \bar{k} = M_d(\bar{k})$, so that $\dim_k D = d^2$. Note that

$$\mathrm{End}_G(V \otimes_k \bar{k}) = D \otimes_k \bar{k} = D \otimes_K (K \otimes_k \bar{k}) \cong D \otimes_K \Big(\prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(K, \bar{k})} \bar{k} \Big) \cong \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(K, \bar{k})} M_d(\bar{k}).$$

This implies, applying semisimplicity over $\bar{k}$, that

$$\rho \otimes_k \bar{k} \cong \bigoplus_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(K,\bar{k})} \rho_\sigma^{\oplus d},$$

where the ρ_σ are pairwise nonisomorphic.

Therefore, on each copy of ρ_τ inside ρ we have:

$$\frac{1}{\#G} \sum_{g \in G} \mathrm{tr}(\rho(g^{-1}A))\rho(g) = \frac{1}{\#G} \sum_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(K,\bar{k})} (d \,\mathrm{tr}_{\bar{k}}\, \rho_\sigma(g^{-1}A))\rho_\tau(g) = d\frac{1}{\dim_{\bar{k}} \rho_\tau} \cdot A,$$

using Schur orthogonality for ρ_τ and the fact that consequently the above sum vanishes unless $\sigma = \tau$. We also used that A induces a well-defined automorphism of each ρ_σ, since $A \in \mathrm{End}_D(V)$ (rather than just $A \in \mathrm{End}_k(V)$).

On the other hand, ρ_σ's all have the same dimension independent of σ, and $[K : k] \cdot d \cdot \dim_{\bar{k}} \rho_\tau = \dim_{\bar{k}}(V \otimes_k \bar{k}) = \dim_k V = [K : k]d^2 \dim_D V$. Therefore $d/(\dim_{\bar{k}} \rho_\tau) = (\dim_D V)^{-1}$ (here we used that $\mathrm{char}\, k = 0$, to ensure that $[K : k] \neq 0$), and the lemma follows. □

Remark 22.43. (i) Now following the proof of Corollary 22.23, etc., one can get a full-fledged 'Fourier inversion package' in the non-algebraically closed case; the main point seems to be to replace each $\dim_k V_i$ with $\dim_{D_i} V_i$.

(ii) The latter sketch of proof shows that there are two factors contributing to $(\rho \otimes_k \bar{k}, V \otimes_k \bar{k})$ reducing: the nontriviality of $[K : k]$, and the nontriviality of the d such that $\dim_k D = d^2$. The former is responsible for there being exactly $[K : k]$-many pairwise nonisomorphic irreducible representations in $\rho \otimes_k \bar{k}$, whereas the latter 'contributes multiplicities of d' to each of these. Only the former occurs for $\mathbb{Z}/3\mathbb{Z}$ acting by rotation on $\mathbb{R}^2$ (which becomes two distinct representations on base-changing to $\mathbb{C}$), and only the latter occurs for the finite quaternion group Q_8 acting on $\mathbb{H}$ (which becomes a sum of two copies of the same representation on base-changing to $\mathbb{C}$).

23. Lecture 23 – Burnside's theorem and Brauer's theorem

23.1. **Burnside's theorem.** In this subsection, we will only consider representations of groups over $k = \mathbb{C}$. All representations will be finite dimensional unless otherwise stated.

Theorem 23.1 (Burnside, 1904)**.** *Any group of order $p^a q^b$, where p and q are prime numbers, is solvable.*

Remark 23.2. While groups of order pqr with p, q and r distinct primes can be shown to be solvable (exercise), A_5 is a group of order $60 = 2^2 \cdot 3 \cdot 5$ which is simple and hence not solvable. As the example of S_3 shows, one cannot replace 'solvable' with 'nilpotent' in Burnside's theorem.

While standard proofs of Sylow's theorems are based on studying groups acting on sets, the proof of Burnside's theorem uses groups acting on vector spaces, i.e., representation theory. Here is one way representations can give us normal subgroups:

Notation 23.3. If (ρ, V) is a representation of G over $\mathbb{C}$, we set

$$Z(V) = \rho^{-1}(Z(GL_{\mathbb{C}}(V))) = \{g \in G \mid \rho(g) = \lambda \cdot \mathrm{Id}_V \text{ for some } \lambda \in \mathbb{C}\}.$$

Clearly, $Z(V) \subset G$ is a normal subgroup.

A key input into the proof of Burnside's theorem is:

Proposition 23.4. *Suppose a finite group G has a conjugacy class $C \neq \{1\}$ of prime power cardinality p^c, p prime and $c \geqslant 0$. Then there exists a nontrivial irreducible representation V of G with $Z(V) \supset C \neq \{1\}$.*

Proof of Burnside's theorem, assuming Proposition 23.4. By induction, we assume that any group H of cardinality strictly dividing $p^a q^b$ is solvable. Therefore, by induction, it is enough to show that G has a nontrivial proper normal subgroup. In particular, we may and do assume that $Z(G) = 1$ – otherwise, either $Z(G) \subset G$ is proper and is the desired normal subgroup, or $G = Z(G)$ is abelian and we are done.

Let $Q \subset G$ be a q-Sylow subgroup. Since Q is a q-group and hence nilpotent, there exists $1 \neq g \in Q$ which centralizes Q (these two are standard facts about nilpotent groups: if you are not comfortable with nilpotent groups, their definition and these two facts are recalled in Remark 23.5 below).

Since g centralizes Q, the cardinality of $C(g)$ is a power of p, and we have $g \neq 1$. Hence by Proposition 23.4, there exists a nontrivial irreducible representation (ρ, V) of G with $Z(V) \neq \{1\}$. If (ρ, V) is not faithful, there is nothing to prove, since $\ker \rho$ is then a proper (by the nontriviality of ρ) and nontrivial subgroup of G.

Thus, assume that ρ is faithful. Then $Z(V)$ the desired nontrivial normal subgroup: note that it is not all of G, since otherwise ρ would inject G into the center of $GL(V)$, forcing G to be abelian, contradicting that $Z(G) = \{1\}$. $\square$

Remark 23.5. For those who are not comfortable with nilpotent groups:

(i) A group G is nilpotent if there exists a chain

$$\{1\} = G_0 \subsetneq G_1 \subsetneq \cdots \subsetneq G_n = G$$

of subgroups of G such that for each $1 \leqslant i \leqslant n$, the commutator $[G, G_i]$ of G and G_i is contained in G_{i-1}. Note that such a group has a nontrivial center: since $[G, G_1] \subset G_0 = \{1\}$, we have that $Z(G) \supset G_1 \supsetneq G_0 = \{1\}$ is nontrivial.

(ii) In the above proof, we used that the group Q, being a q-group, is nilpotent and hence has a nontrivial center. Why is every q-group Q nilpotent? If we show that the center $Z(Q)$ of Q is nontrivial, then we can take $Q_1 = Z(Q) \supsetneq \{1\}$, and induct with Q/Q_1, which is again a q-group. To show that $Z(Q) \neq \{1\}$, use the action of Q on itself by conjugation. All the orbits (i.e., the conjugacy classes) that are not singleton have cardinality a mulitple of q, so, since Q is a q-group, the number of singleton conjugacy classes is a multiple of q. Thus, $\#Z(Q)$ is a multiple of q. But $\#Z(Q) \geqslant 1$, since $1 \in Z(Q)$, so $Z(Q)$ must have at least q elements, and is hence nontrivial.

Now we need to prove Proposition 23.4, for which, in turn, the crucial inputs are the following two lemmas:

Lemma 23.6. *Let G be a finite group. Let $C \subset G$ be a conjugacy class, and (ρ, V) an irreducible representation G, such that $(\#C, \dim V) = 1$. Then either $\chi_\rho(g) = 0$ for $g \in C$, or $C \subset Z(V)$.*

Lemma 23.7. *If G is a finite group with a conjugacy class $C \neq \{1\}$, and p is a prime number, then there exists a nontrivial irreducible representation (ρ, V) of G such that $p \nmid \dim V$, and such that $\chi_\rho(g) \neq 0$ for $g \in C$.*

Proof of Proposition 23.4, assuming Lemmas 23.6 and 23.7. Lemma 23.7 gives us a nontrivial irreducible representation (ρ, V) of G such that $p \nmid \dim V$, and such that $\chi_\rho(g) \neq 0$ for $g \in C$. Note that $(\#C, \dim V) = (p^c, \dim V) = 1$, so Lemma 23.6 applies to this representation (ρ, V) and, since $\chi_\rho(g) \neq 0$ for $g \in C$, forces C to be contained in $Z(V)$. $\square$

It remains to prove Lemmas 23.6 and 23.7. We prove the latter first:

Proof of Lemma 23.7. In this proof, when we write $\sum_V$, it will be understood that the sum is over the set of irreducible representations of G up to isomorphism, and $\sum_{V \neq triv}$ will denote the sub-sum consisting of nontrivial representations. By the Schur orthogonality relations for conjugacy classes, using that $C \neq \{1\}$, we have:

$$0 = \sum_V \chi_V(1^{-1})\chi_V(g) = \sum_V (\dim V)\chi_V(g).$$

Therefore, transferring the contribution of the trivial representation to the other side and dividing by p, we get:

$$\sum_{V \neq triv} \frac{\dim V}{p} \cdot \chi_V(g) = -\frac{1}{p}.$$

Therefore, the left-hand side is not an integer, and hence not an algebraic integer. Since each $\chi_V(g)$, being a sum of its eigenvalues which are $(\#G)$-th roots of unity (since $g^{\#G} = 1$), is an algebraic integer, it follows that there exists a nontrivial irreducible representation V of G such that $p \nmid \dim V$, and $\chi_V(g) \neq 0$. □

To complete the proof of Burnside's theorem (Theorem 23.1), it remains to prove Lemma 23.6. This will in turn need the following input:

Lemma 23.8. *Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are roots of unity, such that*

$$\frac{\varepsilon_1 + \cdots + \varepsilon_n}{n}$$

is an algebraic integer. Then either $\varepsilon_1 + \cdots + \varepsilon_n = 0$, or $\varepsilon_1 = \cdots = \varepsilon_n$.

Let us prove Lemma 23.6 assuming Lemma 23.8:

Proof of Lemma 23.6, assuming Lemma 23.8. Recall from Lecture 22 that $\mathbb{1}_C$ acts on V by an algebraic integer. We also did the following computation in Lecture 22: $\operatorname{tr} \rho(\mathbb{1}_C) = \#C \cdot \chi_\rho(g)$, where g is any element of C, so this algebraic integer is:

$$\#C \cdot \frac{\chi_\rho(g)}{\dim V}.$$

But the following is also an algebraic integer:

$$\chi_\rho(g) = (\dim V) \cdot \frac{\chi_\rho(g)}{\dim V}.$$

Since $\#C$ and $\dim V$ are relatively prime, we can take an integral linear combination of the above two algebraic integers, to get that $\frac{\chi_\rho(g)}{\dim V}$ is an algebraic integer.

But if $n = \dim V$ and $\varepsilon_1, \ldots, \varepsilon_n$ are the eigenvalues of $\rho(g)$, then $(\varepsilon_1 + \cdots + \varepsilon_n)/n = (\chi_\rho(g))/(\dim V)$ is an algebraic integer, so by Lemma 23.8, we have either $\varepsilon_1 + \cdots + \varepsilon_n = 0$ or $\varepsilon_1 = \cdots = \varepsilon_n$. In the former case, $\chi_\rho(g) = 0$ for all $g \in C$, while in the latter case, $\rho(g)$ is a scalar matrix (explanation: since $g^{\#G} = 1$, and since $x^{\#G} - 1$ has distinct roots in $\mathbb{C}$, the minimal polynomial of $\rho(g)$ has distinct roots, so that $\rho(g)$ is diagonalizable over $\mathbb{C}$; being diagonalizable with equal eigenvalues, $\rho(g)$ is a scalar matrix) so that $g \in Z(V)$. □

Proof of Lemma 23.8. This needs some very basic Galois theory, which we will assume, though we have not discussed it yet. Assume that $\varepsilon_1 + \cdots + \varepsilon_n \neq 0$, and that not all the ε_i are equal (to get a contradiction). Then $|(\varepsilon_1 + \cdots + \varepsilon_n)/n| < 1$ by the Cauchy-Schwarz inequality. Let f be the minimal monic polynomial in $\mathbb{Q}[x]$ satisfied by $\alpha := (\varepsilon_1 + \cdots + \varepsilon_n)/n$. By Galois theory, $f(x) = \prod_{\beta \in I}(x - \beta)$, where I is the set of all $\operatorname{Gal}(K/\mathbb{Q})$-conjugates of α,

with $K/\mathbb{Q}$ being any Galois extension containing $\varepsilon_1, \ldots, \varepsilon_n$. Each $\beta \in I$, being a $\mathrm{Gal}(K/\mathbb{Q})$-conjugate of α, is an algebraic integer, so $f(0) = \pm \prod_{\beta \in I} \beta$, being both an algebraic integer and a rational number, is an integer (the same in fact applies to each coefficient of f, so that $f \in \mathbb{Z}[x]$).

Thus, $f(0) \in \mathbb{Z}$. Since α is nonzero, each $\beta \in I$ is nonzero, so that $f(0) = \pm \prod_{\beta \in I} \beta \neq 0$. Therefore, to get a contradiction, it suffices to show that $|f(0)| < 1$. In turn, this follows if we show that $|\beta| < 1$ for each $\beta \in I$. But this is because each such $|\beta|$ is of the form

$$|\sigma(\alpha)| = |\sigma(\varepsilon_1) + \cdots + \sigma(\varepsilon_n)|/n < 1$$

for some $\sigma \in \mathrm{Gal}(K/\mathbb{Q})$; $\sigma(\varepsilon_1), \ldots, \sigma(\varepsilon_n)$ are also roots of unity, not all equal. □

23.2. **Artin's theorem.** In this subsection, we will again work with representations of a finite group G over $k = \mathbb{C}$ (except in Theorem 23.13, where we will allow k to be arbitrary).

Recall that by the complete reducibility of representations of finite groups over $\mathbb{C}$, any element of the representation ring $R_{\mathbb{C}}(G)$ of G (made from $Rep_{\mathbb{C}}(G)$) is uniquely a $\mathbb{Z}$-linear combination $\sum_i n_i[\pi_i]$, where $[\pi_i]$ is the image in $R_{\mathbb{C}}(G)$ of an irreducible representation π_i of G over $\mathbb{C}$. We will abbreviate $R_{\mathbb{C}}(G)$ to $R(G)$. Recall that for each subgroup $H \subset G$, the induction functor Ind_H^G induces an additive map $R(H) \to R(G)$, whose image is an ideal of $R(G)$ (see Corollary 20.31 from Lecture 20).

The aim of this subsection is to prove

Theorem 23.9 (Artin)**.** *If π is a representation of G, then there exist $n_1, \ldots, n_r \in \mathbb{Z}$ and irreducible representations $\pi_1, \ldots, \pi_r$ of cyclic subgroups $H_1, \ldots, H_r$ of G (over $\mathbb{C}$) such that, in $R(G)$ we have:*

$$(\#G) \cdot [\pi] = \sum_i n_i \, \mathrm{Ind}_{H_i}^G [\pi_i].$$

In other words, we have

$$(\#G) \cdot R(G) = \sum_{\substack{H \subset G \\ H \text{ cyclic}}} \mathrm{Ind}_H^G(R(H)).$$

Let us prepare for the proof of this theorem by some general observations concerning the representation ring.

Notation 23.10. Let $\mathcal{C}(G)$ denote the ring of class functions $G \to \mathbb{C}$ (i.e., conjugation invariant maps $G \to \mathbb{C}$), but with its multiplication given by pointwise multiplication rather than convolution.

Consider the assignment, to each finite dimensional representation ρ of G, of its character $\chi_\rho \in \mathcal{C}(G)$. We have:

(i) χ_ρ depends only on the isomorphism class of ρ.

(ii) For any exact sequence

$$0 \to (\rho', V') \to (\rho, V) \to (\rho'', V'') \to 0$$

of representations of G, we have $\chi_\rho = \chi_{\rho'} + \chi_{\rho''}$ (if $0 \to V' \to V \to V'' \to 0$ is an exact sequence of vector spaces over k, and if $A \in \mathrm{End}_k(V)$ restricts to $A' \in \mathrm{End}_k(V')$ and induces $A'' \in \mathrm{End}_k(V'')$, we have $\mathrm{tr}\, A = \mathrm{tr}\, A' + \mathrm{tr}\, A''$).

(iii) If ρ is the trivial representation, χ_ρ is the identity element of $\mathcal{C}(G)$ (since the multiplication on $\mathcal{C}(G)$ is given by pointwise multiplication).

(iv) We have $\chi_{\rho\otimes\rho'} = \chi_\rho \chi_{\rho'}$ (if $A \in \mathrm{End}_k(V)$ and $B \in \mathrm{End}_k(W)$, then $A \otimes B \in \mathrm{End}_k(V \otimes W)$ satisfies $\det(A \otimes B) = (\det A)(\det B)$).

These properties imply that $\rho \mapsto \chi_\rho$ induces a ring homomorphism $R(G) \to \mathcal{C}(G)$.

Lemma 23.11. *This ring homomorphism $R(G) \to \mathcal{C}(G)$ is an injection, and induces an isomorphism $R(G) \otimes_{\mathbb{Z}} \mathbb{C} \to \mathcal{C}(G)$.*

Proof. Since we have seen that the irreducible characters $\chi_\rho \in \mathcal{C}(G)$ of G are linearly independent, where 'irreducible character' means 'character of an irreducible representation', it follows that $R(G) \to \mathcal{C}(G)$ is injective. Since the χ_ρ are a $\mathbb{C}$-basis for $\mathcal{C}(G)$, it follows that $R(G) \otimes_{\mathbb{Z}} \mathbb{C} \to \mathcal{C}(G)$ is a ring isomorphism. □

Notation 23.12. In what follows, we will often identify $R(G)$ with its image in $\mathcal{C}(G)$ under the map induced by $\rho \mapsto \chi_\rho$, and thus also think of $\mathcal{C}(G)$ as $R(G) \otimes_{\mathbb{Z}} \mathbb{C}$.

One of the results we will need for the proof is Mackey's formula for the induced character:

Theorem 23.13. *For this theorem, let k be any field with $(\#G, \mathrm{char}\, k) = 1$, and (ρ, V) an irreducible representation, over k, of a subgroup H of a finite group G. Then for all $g \in G$:*

$$\chi_{\mathrm{Ind}_H^G V}(g) = \frac{1}{\#H} \sum_{\substack{s \in G \\ s^{-1}gs \in H}} \chi_V(s^{-1}gs). \tag{98}$$

Remark 23.14. (i) The formula shows that $\chi_{\mathrm{Ind}_H^G V}(g)$ vanishes unless g is conjugate to an element of H.

(ii) Suppose G is in addition abelian, and that (ρ, V) is given by a homomorphism $\chi : H \to \mathbb{C}^\times$. In this case the formula says:

$$\chi_{\mathrm{Ind}_H^G V}(g) = \begin{cases} \frac{\#G}{\#H} \cdot \chi(g), & \text{if } g \in H, \text{ and} \\ 0, & \text{otherwise.} \end{cases}.$$

Here is a sketch of another way to see it: not necessarily shorter, but probably useful for the picture it conveys. Verify that in this case $\mathrm{Ind}_H^G \chi$ decomposes as $\bigoplus \chi_1$, where χ_1 runs over the set $\mathcal{X}_1$ of characters $G \to \mathbb{C}^\times$ extending χ. Fixing one such character $\chi_{1,0}$, $\chi_1 \mapsto \chi_1 \chi_{1,0}^{-1}$ is a bijection $\mathcal{X}_1 \to \mathrm{Hom}(G/H, \mathbb{C}^\times)$. Now the formula reduces to the fact that the sum of all the homomorphisms $G/H \to \mathbb{C}^\times$ is the function on G/H that equals $(\#G/\#H)$ at the identity, and 0 elsewhere.

Proof of Theorem 23.13. Recall a description of the induced representation from Lecture 20 (Exercise 20.28): $\mathrm{Ind}_H^G V$ is the unique representation of G that contains (a copy of) the representation V of H, and whose underlying vector space is the sum of the g-translates of V as g ranges over any set $[G/H]$ of representatives for G/H: as vector spaces,

$$\mathrm{Ind}_H^G V = \bigoplus_{s \in [G/H]} s \cdot V.$$

Clearly, the action of $g \in G$ permutes the above summands. Hence, if we compute the trace using a basis for $\mathrm{Ind}_H^G V$ formed of the bases for the $s \cdot V$, we get:

$$\chi_{\mathrm{Ind}_H^G V}(g) = \sum_{\{s | g \cdot sH = sH\}} \mathrm{tr}(g; s \cdot V) = \sum_{\{s \in G | s^{-1} g s \in H\}/H} \mathrm{tr}(g; s \cdot V) = \frac{1}{\#H} \sum_{\substack{s \in G \\ s^{-1} g s \in H}} \mathrm{tr}(g; s \cdot V).$$

Now it suffices to observe that $\mathrm{tr}(g; s \cdot V) = \mathrm{tr}(s^{-1} g s; V)$, as one sees from the commutativity of the following diagram, all whose arrows are vector space isomorphisms:

$$\begin{array}{ccc} V & \xrightarrow{s} & s \cdot V \\ {\scriptstyle s^{-1}gs}\uparrow & & \uparrow{\scriptstyle g} \\ V & \xrightarrow{s} & s \cdot V \end{array}.$$

□

Notation 23.15. We have viewed Ind_H^G as a functor $\mathrm{Ind}_H^G : Rep_{\mathbb{C}}(H) \rightsquigarrow Rep_{\mathbb{C}}(G)$ as well as an additive map $\mathrm{Ind}_H^G : R(H) \to R(G)$.

In addition to these two uses of Ind_H^G, we will now add a third: we will also view it as an additive map

$$\mathcal{C}(H) \cong R(H) \otimes_{\mathbb{Z}} \mathbb{C} \to R(G) \otimes_{\mathbb{Z}} \mathbb{C} \cong \mathcal{C}(G).$$

This map then sends each χ_ρ to $\chi_{\mathrm{Ind}_H^G \rho}$, and hence at the level of functions $H \to \mathbb{C}$ and $G \to \mathbb{C}$ is explicitly given by Mackey's formula, (98) from the statement of Theorem 23.13. It restricts to the map $\mathrm{Ind}_H^G : R(H) \to R(G)$.

Proof of Artin's theorem. We identify $R(G)$ with a subspace of $\mathcal{C}(G)$ via $[\pi] \mapsto \chi_\pi$, as in Notation 23.12. Similarly with each $R(H)$.

In this proof, when we write $\sum_H$, the sum will be over the set of cyclic subgroups of G.

Since $\sum_H \mathrm{Ind}_H^G R(H)$ is an ideal of $R(G)$, it is enough to show that it contains $\#G$ times the trivial representation, namely the constant function $g \mapsto \#G$ in $\mathcal{C}(G)$.

For each cyclic subgroup $H \subset G$, define $\chi_H : H \to \mathbb{C}$ by letting $\chi_H(h)$ equal $\#H$ if $\langle h \rangle = H$ (i.e., if h generates H), and 0 otherwise.

This will follow if we prove the following two statements:

(i) Each χ_H belongs to $R(H)$; and

(ii) $\sum_{H\subset G} \mathrm{Ind}_H^G(\chi_H) \in R(G)$, viewed as an element of $\mathcal{C}(G)$, is the constant function $G \to \mathbb{C}$ with value $\#G$.

Let us prove the first of these assertions. By induction, we may assume that $\chi_K \in R(K)$ for each proper subgroup $K \subsetneq H$. Therefore, $\mathrm{Ind}_K^H \chi_K \in R(H)$. It follows from Mackey's formula for induced character (as we saw in Remark 23.14(ii)) that, if $h \in H$ and if $K \subset H$ is a subgroup, then

$$\mathrm{Ind}_K^H \chi_K(h) = \begin{cases} 0, & \text{if } x \notin K, \text{ and} \\ \frac{\#H}{\#K} \cdot \chi_K, & \text{otherwise} \end{cases} = \begin{cases} \#H, & \text{if } \langle h \rangle = K, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

From this, it is easy to see that

$$\chi_H = \#H - \sum_{\substack{K\subsetneq H \\ \text{subgroup}}} \mathrm{Ind}_K^H \chi_K,$$

which belongs to $R(H)$ as we have observed that each $\mathrm{Ind}_K^H \chi_K$ belongs to $R(H)$ (note that $\#H$ also belongs to $R(H)$: it is $\#H$ times the class of the trivial representation of H).

Now let us prove the second of the two assertions above: that $\sum_{H\subset G} \mathrm{Ind}_H^G(\chi_H)$ is the constant function with value G. Using Mackey's formula for the induced character, we get

$$\sum_{H\subset G} \mathrm{Ind}_H^G \chi_H(g) = \sum_{H\subset G} \frac{1}{\#H} \sum_{\substack{s\in G \\ s^{-1}gs\in H}} \chi_H(s^{-1}gs) = \sum_{H\subset G} \sum_{\substack{s\in G \\ \langle s^{-1}gs\rangle = H}} 1,$$

which equals

$$\sum_{s\in G} \sum_{\substack{H\subset G \\ \langle s^{-1}gs\rangle = H}} 1 = \sum_{s\in G} 1 = \#G,$$

as desired. □

23.3. **Brauer's theorem – statement and applications.** Unless otherwise stated, all representations in this subsection will be of finite groups on vector spaces over $\mathbb{C}$.

Definition 23.16. A subgroup $H \subset G$ is called p-elementary if it is a direct product $C \times P$, where $C \subset G$ is a cyclic group, and $P \subset G$ is a p-group. Then H has a unique such decomposition provided we impose the further condition that C has order prime to p.

Theorem 23.17 (Brauer's theorem)**.** *We have*

$$R(G) = \sum_{\substack{H\subset G \\ H \text{ } p\text{-elementary}}} \mathrm{Ind}_H^G R(H).$$

In other words, the image of any finite dimensional representation of G in the representation ring of G is a $\mathbb{Z}$-linear combination of (the images of) representations induced from representations of p-elementary subgroups of G.

This theorem has an important corollary, whose importance we will indicate afterwards in the last two subsections of this lecture.

Corollary 23.18. *For any finite dimensional representation π of G, the image $[\pi]$ of π in $R(G)$ can be written as an integral linear combination $\sum_i n_i[\mathrm{Ind}_{K_i}^G \psi_i]$, where each K_i is a subgroup of G and each $\psi_i : K_i \to \mathbb{C}^\times$ is a homomorphism, i.e., a one-dimensional representation.*

Corollary 23.18 will be an immediate consequence of Theorem 23.17, once we prove that each representation of a nilpotent group H is 'monomial', i.e., given such a representation ρ, there exists a subgroup $K \subset H$ and a one-dimensional representation $\psi : K \to \mathbb{C}^\times$ such that $\rho = \mathrm{Ind}_K^H \psi$. Indeed, combine this fact with the transitivity of induction, i.e., the natural isomorphism between $\mathrm{Ind}_K^G = \mathbb{C}[G] \otimes_{\mathbb{C}[K]} -$ and $\mathrm{Ind}_H^G \circ \mathrm{Ind}_K^H = \mathbb{C}[G] \otimes_{\mathbb{C}[H]} (\mathbb{C}[H] \otimes_{\mathbb{C}[K]} -)$.

Changing notation $H \rightsquigarrow G$, we will prove that this is true more generally for supersolvable groups:

Definition 23.19. A finite group G is said to be supersolvable if there exists a chain

$$\{1\} = G_0 \subsetneq G_1 \subsetneq \cdots \subsetneq G_n$$

of subgroups of G such that each G_i is normal in G, and such that each G_i/G_{i-1} is cyclic (exercise: if this is satisfied we may further assume that each G_i/G_{i-1} is cyclic of prime order).

Note that

$\{p$-elementary groups$\} \subset \{$nilpotent groups$\} \subset \{$supersolvable groups$\} \subset \{$solvable groups$\}$.

Now given Theorem 23.17, Corollary 23.18 will follow from:

Proposition 23.20. *If G is supersolvable, then any irreducible representation π of G over $\mathbb{C}$ is monomial, i.e., there exists a subgroup $H \subset G$ and a character $\psi : H \to \mathbb{C}^\times$ such that $\pi \cong \mathrm{Ind}_H^G \psi$.*

Sketch of proof. Any quotient of G is supersolvable, so by induction, it is easy to reduce to the case where π is faithful. Again, the case where G is abelian is immediate, so assume that G is not abelian.

Since G is supersolvable, it has a normal abelian subgroup A which is not contained in $Z(G)$: to see this note that $\bar{G} := G/Z(G)$ is supersolvable as well, and one can simply take A to be the inverse image of a normal cyclic subgroup of $\bar{G}$.

Since π is faithful, and A is not contained in the center of G, $\pi(A)$ is not contained in the center of $\pi(G)$, so that $\pi|_A$ is not χ-isotypic for any character $\chi : A \to \mathbb{C}^\times$. Let $\chi : A \to \mathbb{C}^\times$ be a character that is contained in $\pi|_A$, and let $V_\chi \subset V$ be the χ-isotypic subspace of $(\pi|_A, V)$. Since $\pi|_A$ is not χ-isotypic, the stabilizer H of V_χ in G is a proper

subgroup of G. Note that $\rho(H)$ preserves V_χ, i.e., ρ restricts to a representation of V_χ on H.

Moreover, it is clear that π, being irreducible, is a sum of the g-translates of the V_χ as g ranges over a set $[G/H]$ of representatives of H in G. It is easy to see that this sum is a direct sum (the translates land in spaces where A acts by distinct characters), so by a description of induced representations given in Lecture 19 (Exercise 20.28), it follows that $\mathrm{Ind}_H^G V_\chi = \pi$. Although V_χ may not be one-dimensional, we are done because we can apply induction to the irreducible representation V_χ of H – V_χ is an irreducible representation of H $\mathrm{Ind}_H^G V_\chi = \pi$ is. Here, we used that H is supersolvable, and also that it is a proper subgroup of G, so that induction can indeed apply to H. □

23.4. **Application of Brauer's theorem to field of definition.** Notice that if $\rho : G \to GL_k(V)$ is a representation of G over a field k, then for any field extension K/k, we get a representation

$$G \xrightarrow{\rho} GL_k(V) \overset{T \mapsto T \otimes \mathrm{id}_K}{\to} GL_K(V \otimes_k K)$$

of G on the K-vector space $V \otimes_k K$. This representation will be denoted $\rho \otimes_k K$.

Definition 23.21. Let K be a field. We say that a representation $\tilde{\rho} : G \to GL_K(\tilde{V})$ of a finite group G over K can be realized over a subfield $k \subset K$, if there exists a representation (ρ, V) of G on a vector space V over k, such that the representation $(\rho \otimes_k K, V \otimes_k K)$ of G on the K-vector space $V \otimes_k K$ is isomorphic to $(\tilde{\rho}, \tilde{V})$.

Equivalently, $\tilde{\rho} : G \to GL_K(\tilde{V})$ can be realized over the subfield $k \subset K$ if and only if there exists a k-vector subspace $V \subset \tilde{V}$, which is G-stable (i.e., $\tilde{\rho}(G)$-stable) and satisfies that the obvious map $V \otimes_k K \to \tilde{V}$ (obtained from the bilinear scalar multiplication map $V \times K \to \tilde{V}$) is an isomorphism.

Corollary 23.22. *Suppose that m is an exponent of the finite group G, i.e., $g^m = 1$ for all $g \in G$. Then:*

(i) The obvious homomorphism $R_{\mathbb{Q}(\zeta_m)}(G) \to R_{\mathbb{C}}(G)$, sending the image $[\pi]$ of a representation π over $\mathbb{Q}(\zeta_m)$ to $[\pi \otimes_{\mathbb{Q}(\zeta_m)} \mathbb{C}]$, is an isomorphism.

(ii) Every representation of G over $\mathbb{C}$ can be realized over $\mathbb{Q}(\zeta_m)$.

To deduce (ii) of the corollary from (i) of the corollary, we will use:

Proposition 23.23. *Suppose $\rho_1 : G \to GL(V_1)$ and $\rho_2 : G \to GL(V_2)$ are nonisomorphic irreducible representations of G over k. Then for any field extension K of k, no irreducible subrepresentation of $(\rho_1 \otimes_k K, V_1 \otimes_k K)$ is isomorphic to an irreducible subrepresentation of $(\rho_2 \otimes_k K, V_2 \otimes_k K)$.*

Proof. Extend $(\rho_1, V_1), (\rho_2, V_2)$ to a sequence $(\rho_1, V_1), \ldots, (\rho_r, V_r)$ of irreducible representations of G over k, up to isomorphism. Then we have a ring isomorphism given by the

action map,

$$\prod_{i=1}^{r} \rho_i : k[G] \to \prod_{i=1}^{r} \mathrm{End}_{D_i}(V_i), \tag{99}$$

where $D_i = \mathrm{End}_G(V_i)$, and this is the decomposition of $k[G]$ into simple rings. Note that this is also $G \times G$-equivariant, if we let the action of (h, g) send $A \in \mathrm{End}_{D_i}(V_i)$ to $\rho_i(h) A \rho_i(g)^{-1}$.

Tensoring over K, we get a ring isomorphism

$$K[G] \to \prod_{i=1}^{r} (\mathrm{End}_{D_i}(V_i) \otimes_k K),$$

again given by an action map (but the rings on the right-hand side are no longer necessarily simple). Therefore, viewing the isomorphism for $K[G]$ analogous to (99) as an identification, in terms of simple rings, we conclude that $(\mathrm{End}_{D_1}(V_1) \otimes_k K)$ and $(\mathrm{End}_{D_2}(V_2) \otimes_k K)$ each identify with products of the form $\prod_i \mathrm{End}_{D_{1,i}}(V_{1,i})$ and $\prod_j \mathrm{End}_{D_{2,j}}(V_{2,j})$, where the $V_{1,i}$ and the $V_{2,i}$ are irreducible representations of G over K.

Clearly, no $V_{1,i}$ is isomorphic to any $V_{2,j}$. It therefore suffices to observe that the $V_{1,i}$ are precisely the irreducible components of $V_1 \otimes_k K$, and that the $V_{2,j}$ are precisely the irreducible components of $V_2 \otimes_k K$: this follows from the fact that each $\mathrm{End}_{D_i}(V_i) \otimes_k K$, as a representation of $G \cong G \times \{1\}$ acting via $g \cdot A = g \circ A$ (which is what makes the action map $G \times \{1\}$-equivariant), is a sum of copies of $V_i \otimes_k K$. □

Proof of Corollary 23.22, assuming Theorem 23.17. By Corollary 23.18, (i) will follow if we show that for each subgroup $H \subset G$ and each homomorphism $\psi : H \to \mathbb{C}^\times$, ψ can be defined over $\mathbb{Q}(\zeta_m)$. But this is the case because H has m as an exponent, so that $\psi(H) \subset \mathbb{Q}(\zeta_m)$, proving (i).

Let us deduce (ii). We will use the following notation: for each representation ρ of G over $\mathbb{Q}(\zeta_m)$, we will abbreviate $\rho \otimes_{\mathbb{Q}(\zeta_m)} \mathbb{C}$ to $\rho_{\mathbb{C}}$, and given $a \in R_{\mathbb{Q}(\zeta_m)}(G)$, we will write $[a]_{\mathbb{C}}$ for its image in $R_{\mathbb{C}}(G)$.

It suffices to show that any irreducible representation $\tilde{\rho}$ of G over $\mathbb{C}$ such that $[\tilde{\rho}] \in R_{\mathbb{C}}(G)$ lies in the image of $R_{\mathbb{Q}(\zeta_m)}(G) \to R_{\mathbb{C}}(G)$, is of the form $\rho_{\mathbb{C}}$ for some (necessarily irreducible) representation ρ of G over $\mathbb{Q}(\zeta_m)$.

Since $[\tilde{\rho}]$ is in the image of $R_{\mathbb{Q}(\zeta_m)}(G)$, we can write $[\tilde{\rho}] = \sum_i n_i [\rho_i]_{\mathbb{C}}$, where each n_i is an integer and each ρ_i is an irreducible representation of G over $\mathbb{Q}(\zeta_m)$. It suffices to show that each n_i is nonnegative: that would imply that, setting $\rho := \bigoplus_i \rho_i^{\oplus n_i}$, we have $[\rho_{\mathbb{C}}] = \sum_i n_i [\rho_i]_{\mathbb{C}} = [\tilde{\rho}]$, so that $\chi_{\tilde{\rho}} = \chi_{\rho_{\mathbb{C}}}$, from which it follows that $\tilde{\rho} \cong \rho_{\mathbb{C}}$, as desired.

Suppose for contradiction that some n_{i_0} is nonnegative. Write $(\rho_i)_{\mathbb{C}} = \bigoplus_j \tilde{\rho}_{i,j}^{m_{i,j}}$, with each $\tilde{\rho}_{i,j}$ irreducible. Then in $R_{\mathbb{C}}(G)$, we can write $[\tilde{\rho}] = \sum_{i,j} n_i m_{i,j} [\tilde{\rho}_{i,j}]$. By Proposition 23.23, $\tilde{\rho}_{i_0,j} \not\cong \tilde{\rho}_{i',j'}$ for any i', j' with $i' \neq i_0$, so that in the expansion $[\tilde{\rho}] = \sum_{i,j} n_i m_{i,j} [\tilde{\rho}_{i,j}]$, $n_{i_0} m_{i_0,j} < 0$ for all the (i_0, j) that occur (which is a nonempty set), giving the desired contradiction. □

23.5. **Motivation for Brauer's theorem from number theory.** This subsection is optional, but recommended to skim over.

Now let us explain what is probably the main number-theoretic motivation for Brauer's theorem, very informally. Recall that the holomorphy at $s = 1$ of the Dirichlet L-functions $L(s, \chi)$, associated to Dirichlet characters $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times = GL_1(\mathbb{C})$ (with N varying) was crucial in the proof of Dirichlet's theorem on arithmetic progressions.

Now a Dirichlet character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ can be viewed as a special case of a representation of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$: if ζ_N is a primitive N-th root of unity, then one can show that $\mathbb{Q}(\zeta_N)/\mathbb{Q}$ is Galois, and that sending $\sigma \in \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ to the uniquely defined $l \in \mathbb{Z}/N\mathbb{Z}$ such that $\sigma(\zeta_N) = \zeta_N^l$, defines an isomorphism

$$\mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q}) \to (\mathbb{Z}/N\mathbb{Z})^\times, \tag{100}$$

giving a 'Galois representation' (by which we mean the composite)

$$\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q}) \to (\mathbb{Z}/N\mathbb{Z})^\times \xrightarrow{\chi} \mathbb{C}^\times.$$

Note that the isomorphism (100) does not depend on the choice of ζ_N, since any two choices of ζ_N are powers of each other. The main point of the proof of (100) is the irreducibility of the cyclotomic polynomial associated to N over $\mathbb{Q}$.

Artin's L-function is defined in such a way that if ρ is a one-dimensional representation associated to a Dirichlet character χ as above, then $L(s, \rho) = L(s, \chi)$.

The precise definition of $L(s, \rho)$, for a representation $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Gal}(K/\mathbb{Q}) \to GL_N(\mathbb{C})$ with $K/\mathbb{Q}$ finite Galois, involves number theory, and is only given as an Euler product over primes, that converges in some $\mathrm{Re}\ s \gg 0$.

Question: How do we complete this definition by at least continuing $L(s, \rho)$ meromorphically (in the complex variable s) to $\mathbb{C}$?

This was perhaps Artin's main motivation. We will explain more below.

In fact, to generalize the result that $L(s, \chi)$ is holomorphic on $\mathbb{C}$ for nontrivial χ, Artin conjectured:

Conjecture 23.24. If $\rho : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to GL_m(\mathbb{C})$ is an irreducible Galois representation factoring through $\mathrm{Gal}(K/\mathbb{Q})$ for some finite Galois extension $K/\mathbb{Q}$, and if ρ is nontrivial, then $L(s, \rho)$ analytically continues to a holomorphic function on $\mathbb{C}$.

But how does one either prove such a result or answer the question mentioned above (about meromorphic continuation), especially when $L(s, \rho)$ is only defined in some 'sufficiently right' half plane as an Euler product?

We already know that the conjecture is true for those ρ associated to a Dirichlet character χ. This takes care of the case where ρ is one-dimensional. This can be pushed a bit further, and one can show the following: if $\rho : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_m(\mathbb{C})$ factors through $\bar{\rho} : \mathrm{Gal}(K/\mathbb{Q}) \to \mathrm{GL}_m(\mathbb{C})$, and if $\bar{\rho}$ is of the form $\mathrm{Ind}_{\mathrm{Gal}(K/K_0)}^{\mathrm{Gal}(K/\mathbb{Q})} \xi$, where $\xi : \mathrm{Gal}(K/K_0) \to \mathbb{C}^\times$ is a homomorphism, then one can show that the conjecture is true for ρ, i.e., $L(s, \rho)$ extends

to an analytic function on $\mathbb{C}$. Namely, this involves a slight generalization of the definition of $L(s, \chi)$ with $\mathbb{Q}$ replaced by the number field K_0, where in place of of a Dirichlet character one considers its generalization called a Hecke character, associated to $\xi : \mathrm{Gal}(K/\mathbb{Q}) \to \mathbb{C}^\times$ using class field theory ((100) is also part of class field theory, and is being generalized here).

Write $G = \mathrm{Gal}(K/\mathbb{Q})$, and consider $L(s, \rho)$, where ρ varies over representations of $\mathrm{Gal}(K/\mathbb{Q})$, attached to ρ (pulled back to $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$). The above discussion says that we know the analytic continuation for $L(s, \rho)$ when ρ is of the form $\mathrm{Ind}_H^G \xi$, where $H \subset G$, and $\xi : H \to \mathbb{C}^\times$ is a character.

Moreover, the definition of $L(s, \rho)$ is such that $L(s, \rho_1 \oplus \rho_2) = L(s, \rho_1)L(s, \rho_2)$. Thus, one can get *meromorphic continuation* for $L(s, \rho)$ if one knows that

$$\rho = \bigoplus_i \rho_i^{\oplus n_i},$$

where each ρ_i is of the form $\mathrm{Ind}_{H_i}^G \xi_i$, and each $n_i \in \mathbb{Z}$ (since the meromorphic continuation is known for each $L(s, \rho_i)$). This is precisely what Brauer's theorem does. This completes our description of the number theoretic motivation for Brauer's theorem.

Note that this only gives meromorphic continuation, and does not imply that $L(s, \rho)$ is analytic for nontrivial irreducible ρ. In fact, this assertion is still open; many special cases that have been proved are spectacular results, but these in a sense constitute a tiny proportion of the total number of cases. It seems to be generally understood that to prove such an analyticity, like in the dimension one case, one will need to attach to $\rho : \mathrm{Gal}(K/\mathbb{Q}) \to GL_m(\mathbb{C})$ an object generalizing a Dirichlet character.

(Lots of caveats apply to all of the following). When $m = 2$ an object of this sort is a modular form, for larger m such an object is what is known as an automorphic representation, etc. The point is: though a modular form might appear complicated, one can attach an L-function to a modular form (just like to a Dirichlet character), and prove its analytic continuation, which is very difficult to do for Artin's $L(s, \rho)$.

Thus, one wants to see if one can get a 'correspondence' between 'Galois representations' and 'automorphic representations'. This is the subject of study of an important and influential area of mathematics called the Langlands program, named after Robert Langlands.

24. Lecture 24 – Algebraic closure, separable extensions

Today, unless otherwise specified, k, K, L etc. will denote fields.

24.1. **Existence of algebraic closure.** Recall that a field extension of a field k is a k-algebra that is a field; e.g., $k(t) \to k(t)$, $t \mapsto t^2$, is a field extension. Recall that any homomorphism of fields is injective. We will use the following facts from field theory, so we leave them as exercises (similar but more basic facts may be implicitly assumed):

Exercise 24.1. (i) If K/k is an algebraic field extension, then any k-algebra homomorphism $K \to K$ is an isomorphism.
Hint/note: This is immediate if K is generated by a single element over k (i.e., as a k-algebra). The general case can be reduced easily to this. See Lemma 2.1, Chapter V (Algebraic extensions) of Serge Lang's book.

(ii) A splitting field of any *finite* family $\{f_j \mid j \in J\} \subset k[x]$ of polynomials in $k[x]$ is unique up to a usually non-unique isomorphism, and embeds into any other extension in which f splits into linear factors (the infinite case will be considered below).

Remark 24.2. The zero ring is a k-algebra. It is the unique final object in the category of k-algebras, k being the unique initial object.

We first prove Artin's theorem that any field k has an algebraic closure:

Theorem 24.3. *Any field k has an algebraic closure, i.e., there exists an extension $k \hookrightarrow L$ such that L/k is algebraic, and such that L is algebraically closed.*

Proposition 24.4. *If k is a field and $I \subset k[x]$ is an arbitrary family of polynomials in $k[x]$, then this family is split by some field: there exists a field extension L/k such that each $f \in I$ splits into linear factors in $L[x]$ (constant polynomials are assumed by convention to be already split into linear factors).*

Proof that Proposition 24.4 implies Theorem 24.3. Let $I = k[x]$, and choose L as in Proposition 24.4. The set of elements of L that are algebraic over k form a subfield $L_0 \subset L$, and all polynomials in $k[x]$ clearly split completely in L_0, so we may replace L with L_0 and assume that L is algebraic over k. It is enough to prove that L is algebraically closed.

If not, it has a finite nontrivial algebraic extension $L[\alpha]$. Then α is algebraic over L and hence over k, and is therefore a root of an irreducible polynomial $f \in k[x]$. However, since L splits f, this forces $\alpha \in L$, a contradiction to $L[\alpha]/L$ being nontrivial. □

Proof of Proposition 24.4. The main ideas in the two proofs of this theorem that we will outline/discuss seem to be:

(i) Any finite collection of polynomials in $k[x]$ has a splitting field; and
(ii) Using Zorn's lemma, this can be extended to infinite collections.

A crude outline of one proof is as follows: any algebraic extension of k can be shown to have cardinality at most $\max(\aleph_0, |k|)$, where $\aleph_0$ is the cardinality of the set of natural numbers, so choose a set S of bigger cardinality, and consider the collection of all algebraic extensions $k \hookrightarrow K$ such that the underlying set of K is a subset of S. These form a set, and one can use a Zorn's lemma argument. For more details, see https://stacks.math.columbia.edu/tag/09GP

Now we discuss another proof, which I have taken from a post by Tom Leinster in n-category cafe (https://golem.ph.utexas.edu/category/2021/04/algebraic_closure.html), which is based on a proof expounded by Keith Conrad, in turn based on an argument by Zorn in the paper where he first introduced Zorn's lemma.

It is enough to show that there exists a *nonzero* k-algebra $k \hookrightarrow S$, such that each polynomial in I splits into linear factors in S. Indeed, given such an S, let $\mathfrak{m} \subset S$ be a maximal ideal, and set $L := S/\mathfrak{m}$ (note that the existence of such an $\mathfrak{m}$ depends on Zorn's lemma). Then $k \to L$ is an algebraic extension in which each element of I splits into linear factors.

Now let us construct such an S. For each $f \in k[x]$ of degree at least 1, we can not only construct a canonical field extension containing a root of f, but also a "splitting ring extension" that splits f completely:

$$SR_k(f) = k[\alpha_{f,1}, \ldots, \alpha_{f,\deg f}]/J_f,$$

where $k[\alpha_{f,1}, \ldots, \alpha_{f,\deg f}]$ is a polynomial ring in $\deg f$-many variables, and $J_f \subset k[\alpha_{f,1}, \ldots, \alpha_{f,\deg f}]$ is the ideal generated by the coefficients of:

$$f(x) - a \prod_{i=1}^{\deg f} (x - \alpha_{f,i}),$$

where a is the leading coefficient of f. If f is constant, set $SR_k(f) = k$.

One can similarly define $SR_k(I')$ for each subset $I' \subset k[x]$:

$$SR_k(I') = k[\{\alpha_{f,i} \mid f \in I', 1 \leqslant i \leqslant \deg f\}]/J_{I'}, \tag{101}$$

where $J_{I'}$ is the ideal generated by the J_f as f varies over I' (note that when I' is finite, $SR_k(I')$ is the coproduct of the $SR_k(f)$ as f varies over I', in the category of commutative k-algebras). Here, for those $f \in I'$ which are constant, by convention there are no $\alpha_{f,i}$, and no contribution to $J_{I'}$.

We would be done if we could show that $SR_k(I)$ is nonzero, but this does not seem easy. However, it is easy to see that $SR_k(I')$ is nonzero for each finite subset $I' \subset k[x]$: this is because there is a finite extension K/k in which each $f \in I'$ splits into linear factors, and then the map $k[\{\alpha_{f,i} \mid f \in I', 1 \leqslant i \leqslant \deg f\}] \to K$ sending, for each $f \in I'$, $\alpha_{f,1}, \ldots, \alpha_{f,\deg f}$ to the roots of f in K arranged in some order, vanishes on $J_{I'}$. Hence it factors through a ring homomorphism $SR_k(I') \to K$ that sends 1 to 1, and is hence nonzero (since $1 \neq 0$ in K), forcing $SR_k(I') \neq 0$.

One now considers the collection of finite subsets $I' \subset k[x]$, and make them into a directed set $(\mathcal{I}, \leqslant)$ under inclusion. There is an obvious map $SR_k(I') \to SR_k(I'')$ whenever $I \subset I'$,

induced by sending each $\alpha_{f,i}$ to $\alpha_{f,i}$, and this gives a directed system of rings indexed by $\mathcal{I}$. Set:

$$S := \varinjlim_{I'} SR_k(I').$$

The point is that S is a directed colimit of nonzero rings, and is hence nonzero: this is because otherwise we get $0 = 1$ in S, which would force the same to be the case in $SR_K(I')$ for some finite set I', a contradiction. Clearly any nonconstant polynomial in $k[x]$ factors into linear polynomials in S, as desired. □

Proposition 24.5. *Let K/k be an algebraic extension.*

(i) If $k \hookrightarrow L$ is any field extension with L an algebraically closed field, there exists a k-algebra embedding $K \hookrightarrow L$.
(ii) In the situation of (i), if further K is algebraically closed and L is an algebraic extension of K, then any embedding $\sigma : K \hookrightarrow L$ as in (i) is an isomorphism.
(iii) An algebraic closure of k is unique up to an isomorphism (though usually not up to a unique isomorphism).

Proof. First let us sketch a proof of (i). If $K = k[\alpha]$, where the minimal polynomial of α is f, then we can map $\alpha \in K$ to any root of f in L. The general case of (i) is an easy Zorn's lemma argument based on this; see Theorem 2.8 in Serge Lang's book.

In the situation of (ii), note that σ is automatically injective (being a homomorphism of fields), but also surjective since $\sigma(K)$ is algebraically closed and L is algebraic over $\sigma(K)$. (iii) follows from (ii). □

Now let us study the non-uniqueness of the isomorphism up to which an algebraic closure is unique:

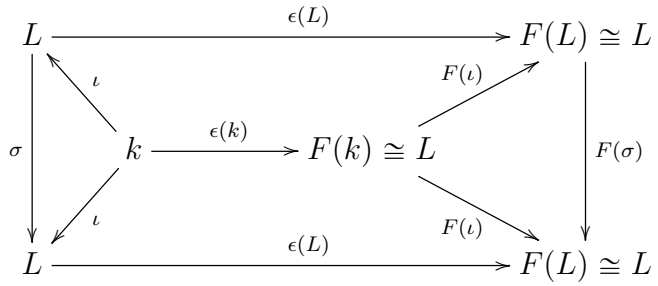
Lemma 24.6. *Let Fld be the category of fields, $AlgClosFld$ the full subcategory of algebraically closed fields, and $G : AlgClosFld \rightsquigarrow Fld$ the obvious inclusion functor.*

(i) Algebraic closure cannot be defined functorially, i.e., there is no functor $F : Fld \rightsquigarrow AlgClosFld$ together with a natural transformation $\epsilon : \mathrm{id}_{Fld} \rightsquigarrow G \circ F$, with the property that for all fields k, $\epsilon(k) : k \to G \circ F(k)$ is an algebraic closure of k.
(ii) The subcategory $AlgClosFld$ of Fld is not a reflective subcategory, i.e., there is no left adjoint $F : Fld \rightsquigarrow AlgClosFld$ to $G : AlgClosFld \rightsquigarrow Fld$.

Remark 24.7. The condition in (i) of the lemma is exactly the functoriality of the algebraic closure: the datum of an algebraic closure of k is not just an algebraically closed field $\bar{k}$, but comes with the datum of a field extension $k \hookrightarrow \bar{k}$ as well. The functoriality of the algebraic closure refers to defining the algebraic closure at the level of morphisms as well, in such a manner that $k \hookrightarrow \bar{k}$ is functorial in k; in otherwords, it includes a natural transformation from the identity functor $\mathrm{id}_{Fld} : k \rightsquigarrow k$ to the putative algebraic closure functor $k \rightsquigarrow \bar{k}$.

Proof of Lemma 24.6. Let us prove (i). The reason for the failure of the condition in (i) will turn out to be the fact that an algebraic closure typically has nontrivial automorphisms over the base field, so fix any pair $(\iota : k \hookrightarrow L, \sigma)$ consisting of an algebraic closure $\iota : k \hookrightarrow L$ of a field k, and a *nontrivial* k-algebra automorphism $\sigma : L \to L$. For instance $\iota : k \hookrightarrow L$ could be the inclusion $\mathbb{R} \hookrightarrow \mathbb{C}$, and σ could be the complex conjugation.

If such a functor F existed, we would then get a commutative diagram as follows:



Here, the commutativity of the rectangle and the two trapezia is a consequence of the naturality of ϵ. Note that, by definition, each horizontal arrow in the above diagram is an algebraic closure. The commutativity of the rectangle in the above diagram, together with the fact that σ is nontrivial, would force $F(\sigma)$ to not be the identity map $F(L) \to F(L)$, while the commutativity of the right triangle would force $F(\sigma)$ to be the identity, a contradiction.

This proves (i), and we come to (ii). If such a left-adjoint F existed, then since G is fully faithful, problem 1 of HW 3 would imply that the counit of an adjunction between F and G would be a natural isomorphism $\epsilon : F \circ G \rightsquigarrow \mathrm{id}$. In particular, $F(L) = F \circ G(L)$ would be isomorphic to L if L is algebraically closed. Now consider, for a finite field $\mathbb{F}_q$ and an algebraic closure $\mathbb{F}_q \hookrightarrow \bar{\mathbb{F}}_q$ of it, the chain of bijections:

$$\mathrm{Hom}_{Fld}(\mathbb{F}_q, \bar{\mathbb{F}}_q) \overset{\text{via } \bar{\mathbb{F}}_q \cong G(\bar{\mathbb{F}}_q)}{\to} \mathrm{Hom}_{Fld}(\mathbb{F}_q, G(\bar{\mathbb{F}}_q)) \to \mathrm{Hom}_{Fld}(F(\mathbb{F}_q), \bar{\mathbb{F}}_q).$$

The left-hand side is clearly finite, since any homomorphism $\mathbb{F}_q \hookrightarrow \bar{\mathbb{F}}_q$ has image in the copy of $\mathbb{F}_q$ inside $\bar{\mathbb{F}}_q$ (since this copy is defined by $x^q = x$), while it is easy to see that $F(\mathbb{F}_q)$, being an algebraically closed field contained in $F(\bar{\mathbb{F}}_q) \cong \bar{\mathbb{F}}_q$, is isomorphic to $\bar{\mathbb{F}}_q$, so the right-hand side is not finite. □

Remark 24.8. The condition defining a reflective subcategory, that of its inclusion having a left adjoint, is something we have seen often: *AbGrp* is a reflective subcategory of *Grp*, since the inclusion functor *AbGrp* $\rightsquigarrow$ *Grp* has abelianization as a left adjoint. Another example is the inclusion of the category of compact Hausdorff topological spaces in the category of all topological spaces, which has the Stone-Čech compactification as a left adjoint.

24.2. **Separable degree.** The uniqueness of algebraic closure, even though up to a non-unique isomorphism, allows us to define the notion of separable degree.

Corollary 24.9. *If K/k is an algebraic extension, then for any two algebraically closed field extensions L_1, L_2 of k, there is a bijection*

$$\mathrm{Hom}_{k\text{-}Alg}(K, L_1) \to \mathrm{Hom}_{k\text{-}Alg}(K, L_2).$$

Proof. Without loss of generality, we may and do assume that L_1 is an algebraic closure of k. Then by Proposition 24.5(i), there exists a k-algebra embedding $\sigma : L_1 \hookrightarrow L_2$. Now $\sigma(L_1) \subset L_2$ is an algebraically closed field containing k, so any k-algebra homomorphism $K \to L_2$ factors through $\sigma(L_1)$, and hence through L_1. Therefore, composing with σ gives the required bijection. □

We would like to study $k\text{-}Alg^{fc}$, the category of finite commutative k-algebras, where finite means finite dimensional as a vector space over k (and in particular Artinian). Since we are going to frequently encounter the following notation, we state it separately:

Notation 24.10. We will frequently use the following notation.

(i) If $A \in \mathrm{Ob}\, k\text{-}Alg^{fc}$, then by the structure theory for commutative Artinian rings, we have a decomposition
$$A = \prod_{i=1}^{r} A_i,$$
with each A_i Artinian local.

(ii) Note that each A_i as above has an obvious structure of a k-algebra: $k \to A \to A_i$, something that will be used without mention from now on.

(iii) For $1 \leqslant i \leqslant r$, if $\mathfrak{m}_i = \mathrm{rad}(A_i)$ is the radical – or in other words the unique maximal ideal – of the Artin local ring A_i, then $\mathfrak{m}_i \subset A_i$ is nilpotent (the radical of an Artinian ring being nilpotent), and the residue field $K_i := A_i/\mathfrak{m}_i$ is a finite extension of k, via
$$k \hookrightarrow A_i \to A_i/\mathfrak{m}_i = K_i.$$

Exercise 24.11. Show that, in the setting of Notation 24.10, the map $A \to \prod_{i=1}^r A_i \to \prod_{i=1}^r K_i$ induces isomorphisms

$$A/\operatorname{rad}(A) \to \prod_{i=1}^{r} (A_i/\operatorname{rad}(A_i)) = \prod_{i=1}^{r} (A_i/\mathfrak{m}_i) \to \prod_{i=1}^{r} K_i.$$

In what follows, we will use this frequently as well.

Remark 24.12. If k is algebraically closed, each inclusion $k \hookrightarrow K_i$ is an isomorphism (since $\dim_k K_i < \infty$), and hence

$$A/\operatorname{rad}(A) \cong \prod_{i=1}^{r} K_i \cong \prod_{i=1}^{r} k \tag{102}$$

is a product of copies of k. This will be used often in what follows as well.

Corollary 24.13. *If A is a finite commutative k-algebra, then there exists a unique natural number $[A : k]_s$ such that for any k-algebra $k \hookrightarrow L$ with L an algebraically closed field,*

$$[A : k]_s = \#\operatorname{Hom}_{k\text{-}Alg}(A, L).$$

Proof. For any two algebraically closed field extensions L_1, L_2 of k, the proof of Corollary 24.9 goes through to give a bijection $\operatorname{Hom}_{k\text{-}Alg}(A, L_1) \to \operatorname{Hom}_{k\text{-}Alg}(A, L_2)$ – use that the image of any k-algebra homomorphism $A \to L_2$ is an integral domain (being contained in L_2) which is finite dimensional over k (since $\dim_k A < \infty$), and is hence a field. Therefore, the right-hand side is independent of the choice of L.

It remains to see that it is finite; let us do it in a manner that will also give us a picture of how elements of $\operatorname{Hom}_{k\text{-}Alg}(A, L)$ look like. Associate to A the A_i and the K_i, $1 \leqslant i \leqslant r$, as in the above discussion. Then each k-algebra homomorphism $A \to L$ factors through $A \to A_i$ for a unique $1 \leqslant i \leqslant r$ (use that $A_i A_j = 0$ in A for $i \neq j$), and then through $A \to K_i$ (since $\mathfrak{m}_i = \operatorname{rad}(A_i)$ is nilpotent), giving us a bijection

$$\operatorname{Hom}_{k\text{-}Alg}(A, L) \to \bigsqcup_{i=1}^{r} \operatorname{Hom}_{k\text{-}Alg}(K_i, L).$$

Now note that for each $1 \leqslant i \leqslant r$, $\operatorname{Hom}_{k\text{-}Alg}(K_i, L)$ is finite as $\dim_k K_i \leqslant \dim_k A_i \leqslant \dim_k A < \infty$. □

Definition 24.14.
(i) For a finite commutative k-algebra A, the number $[A : k]_s$ as in Corollary 24.13 will be called the separable degree of A over k.
(ii) For each algebraic (possibly infinite) field extension K of k, we similarly have $[K : k]_s := \#\operatorname{Hom}_{k\text{-}Alg}(K, L) \in \mathbb{N}_{\geqslant 1} \cup \{\infty\}$ independent of the choice of an algebraically closed field L containing k. Since K/k may be finite, we are allowing the possibility $[K : k]_s = \infty$.

Remark 24.15. It is possible that a more correct approach would be to unify the above two definitions by defining $[A : k]_s$ for a class of A that includes both finite commutative k-algebras and possibly infinite algebraic field extensions. Possibly, the 'correct category' to look at is that of k-algebras A that are directed colimits of finite commutative k-algebras. However, right now I don't have the time to figure out the correct thing.

Notation 24.16. For a finite commutative k-algebra A, we will write $[A : k]$ for $\dim_k A$. Thus, $[A : k]$ generalizes the notion of the degree of a field extension. (Like with fields, we will see that we have $[A : k]_s \leqslant [A : k]$, with equality defining the notion of separability.) Similarly, we have $[K : k] = \dim_k K \in \mathbb{N}_{\geqslant 1} \cup \{\infty\}$ for an algebraic field extension K/k.

Example 24.17. If k is algebraically closed, and $A = \prod_{i=1}^{r} A_i$ is a finite commutative k-algebra, with each A_i local, then we claim:

$$[A : k]_s = \dim_k A/(\operatorname{rad} A) = r \leqslant \dim_k A = [A : k].$$

This follows from the fact that, by Remark 24.12, we have for any algebraically closed field L containing k (please work out all the steps below as an exercise):

$$[A:k]_s = \#\operatorname{Hom}_{k\text{-}Alg}(A,L) = \#\operatorname{Hom}_{k\text{-}Alg}(A/\operatorname{rad}(A),L) = \#\operatorname{Hom}_{k\text{-}Alg}\Big(\prod_{i=1}^{r} k, L\Big) = r \leqslant \dim_k A = [A:k].$$

Proposition 24.18. *If $K/E/k$ is a chain of field extensions, then (with the understanding that $n \cdot \infty = \infty \cdot n = \infty$ for all $n \in \mathbb{N}_{\geqslant 1} \cup \{\infty\}$):*

$$[K:k]_s = [K:E]_s \cdot [E:k]_s.$$

Proof. Let $k \hookrightarrow L$ be a k-algebra with L an algebraically closed field. By Proposition 24.5(i), restriction from K to E induces a surjection

$$\operatorname{Hom}_{k\text{-}Alg}(K,L) \to \operatorname{Hom}_{k\text{-}Alg}(E,L).$$

Since these sets have cardinalities $[K:k]_s$ and $[E:k]_s$, it is enough to show that the fiber of the above map over each $\varphi \in \operatorname{Hom}_{k\text{-}Alg}(E,L)$ has cardinality $[K:E]_s$. But this is so by the definition of $[K:E]_s$, because this fiber is precisely $\operatorname{Hom}_{E\text{-}Alg}(K,L)$, where L is thought of as an E-algebra via $\varphi : E \hookrightarrow L$ (thus, the point is that this fiber has cardinality $[K:E]_s$ independently of $\varphi : E \hookrightarrow L$, by Corollary 24.9). □

Exercise 24.19. Formulate an analogue of Proposition 24.18 that applies with K and E replaced by finite commutative k-algebras A and A_1. Note that the formulation will necessarily be less simple than in the proposition.

One motivation for studying separable degree is the following easy lemma:

Lemma 24.20. *If $K = k[\alpha]$ is a finite algebraic field extension of k generated by a single element α with minimal polynomial, say $f \in k[x]$, then for any algebraically closed field L containing k, $[K:k]_s$ is the number of distinct roots of f in L.*

Proof. Immediate from basic field theory. □

The assertion in the following exercise, or equivalently (103), will be used often in what follows.

Exercise 24.21. If $R \to S$ is a morphism of rings, recall the Hom-tensor adjointness isomorphism

$$\operatorname{Hom}_S(S \otimes_R M, N) \to \operatorname{Hom}_R(M,N),$$

for all left R-modules M and left S-modules N, given by composition with the map $M \to S \otimes_R M$ sending each m to $1 \otimes m$. When R and S are commutative, M is an R-algebra and N is an S-algebra, verify that the above adjunction isomorphism restricts to a bijection:

$$\operatorname{Hom}_{S\text{-}Alg}(S \otimes_R M, N) \to \operatorname{Hom}_{R\text{-}Alg}(M,N). \tag{103}$$

In particular, the extension of scalars functor $S \otimes_R - : R\text{-}Alg \rightsquigarrow S\text{-}Alg$ is left adjoint to the restriction of scalars functor $S\text{-}Alg \rightsquigarrow R\text{-}Alg$.

Proposition 24.22. *(i) If A/k is a finite commutative k-algebra, and F/k is a (not necessarily algebraic) field extension, then $[A : k]_s = [F \otimes_k A : F]_s$ (sanity check: $F \otimes_k A$ is a finite commutative F-algebra).*

(ii) If k is algebraically closed, then $[A : k]_s = [A/\operatorname{rad}(A) : k] \leqslant [A : k]$.

(iii) We have $[A : k]_s \leqslant [A : k] := \dim_k A$. Similarly, $[K : k]_s \leqslant [K : k]$ for any algebraic field extension K/k.

Remark 24.23. An assertion like (i) seems difficult to formulate if we work with only field extensions and not finite commutative k-algebras; this is because fields are not closed under tensor product. Thus, working with finite commutative k-algebras seems to give us better flexibility.

Proof of Proposition 24.22. For (i), let L/F be an extension with L algebraically closed, so we have $k \hookrightarrow F \hookrightarrow L$, which we use to compute $[A : k]_s$. Then by (103), with k, F, A and L in place of R, S, M and N, we have a bijection

$$\operatorname{Hom}_{k\text{-}Alg}(A, L) \to \operatorname{Hom}_{F\text{-}Alg}(A \otimes_k F, L),$$

which gives $[A : k]_s = [A \otimes_k F : F]_s$. Thus, (i) holds.

(ii) has already been worked out in Example 24.17. To see (iii) for finite commutative k-algebras – we leave it as an exercise to deduce the field extension case from there – letting L be an algebraic closure of k, we have

$$[A : k]_s \overset{\text{(i)}}{=} [A \otimes_k L : L]_s \overset{\text{(ii)}}{\leqslant} [A \otimes_k L : L] = [A : k].$$

□

24.3. **Separable algebras.**

Definition 24.24. (i) A finite commutative k-algebra A is said to be separable or étale over k if the inequality $[A : k]_s \leqslant [A : k]$ from Proposition 24.22(iii) is an equality.

(ii) If K/k is a field extension and $\alpha \in K$ is algebraic over k, we say that α is separable over k if $k[\alpha]/k$ is a separable field extension.

(iii) If K/k is an infinite algebraic field extension, we say that K/k is separable if each finite subextension of K/k is. Note that, a priori, this definition applies only to infinite extensions. Though this is unsatisfactory in that the definitions for finite and infinite extensions seem very different from each other, it will follow from Corollary 24.33 below (see Remark 24.34) that an equivalent finiteness-agnostic definition for separability of field extensions can be given.

(iv) A finite extension K/k of fields is called purely inseparable if $[K : k]_s = 1$. An infinite extension K/k of fields is called purely inseparable if every finite subextension of K/k is purely inseparable.

(v) If $R \to S$ is a homomorphism of commutative rings, such that S is a finite free R-module, we will write $\operatorname{tr}_{S/R} : S \to R$ (resp., $N_{S/R} : S \to R$) for the map that sends each $a \in S$ to the trace (resp., the determinant) of the 'multiplication by a'

map $(a \mapsto sa) \in \mathrm{End}_R(S)$. (Exercise: make sense of $\mathrm{tr}_{S/R}$ when S is only finite projective over R). Note that $\mathrm{tr}_{S/R} : S \to R$ is a homomorphism of additive groups, while $N_{S/R} : S \to R$ is a homomorphism of multiplicative monoids.

(vi) Further, if R and S are as above, we use the same notation $\mathrm{tr}_{S/R}$ to also denote the (clearly symmetric) bilinear form:

$$\mathrm{tr}_{S/R} : S \times S \to S,$$

that sends $(a, b) \in S \times S$ to $\mathrm{tr}_{S/R}(ab)$.

The definition of $\mathrm{tr}_{S/R}$ will be of interest to us mainly when R is a field k and A is a finite commutative algebra over it.

The following lemma gives us a feel for separability of elements:

Lemma 24.25. *If K/k is a field extension and $\alpha \in K\backslash\{0\}$ is algebraic over k, then the following are equivalent:*

(i) α is separable over k.
(ii) The minimal polynomial $f \in k[x]$ of α is separable, i.e., f has distinct roots in L for some or equivalently any algebraically closed field L containing k.
(iii) $f' \neq 0$.

These conditions are violated if and only if the following hold: $\mathrm{char}\, k$ *is some $p > 0$, and we can write $f(x) = g(x^p)$ for some $g \in k[x]$.*

Proof. The equivalence of (i) and (ii) is an immediate consequence of the definition and Lemma 24.20. Let us show that (ii) is equivalent to (iii). f has a repeated root if and only if α itself is a repeated root of f, which is the case if and only if $f'(\alpha) = 0$: use Leibniz' product rule, which gives that if $f(x) = (x - \alpha_1) \dots (x - \alpha_r)$ with $\alpha = \alpha_1$, then $f'(\alpha) = \prod(\alpha - \alpha_2) \dots (\alpha - \alpha_r)$. Since f is a minimal polynomial for α and $\deg f' < \deg f$, this is the case if and only if $f' = 0$; in other words, we have (ii) $\iff$ (iii).

For the last assertion, note that f' never vanishes in characteristic zero (the minimal polynomial of $\alpha \in K\backslash\{0\}$ is nonconstant), and that when $\mathrm{char}\, k = p > 0$, f' vanishes if and only if f is a polynomial in x^p, i.e., if and only if $f(x) = g(x^p)$ for some $g \in k[x]$. □

Example 24.26. Thus, it follows that if k is of characteristic zero, any algebraic field extension of k is separable. On the other hand, if k is a field of characteristic p and if there exists $a \in k^\times\backslash(k^\times)^p$, then $f(x) = x^p - a$ is irreducible, and adjoining a root α of f gives an extension $k[\alpha]/k$ which is purely inseparable, since f is a polynomial in x^p.

In particular, when $\mathrm{char}\, k = p > 0$, the extension $k(t) \to k(t)$ given by $t \mapsto t^p$ is not separable (this example recalls that field extensions are really homomorphisms and not necessarily "physical inclusions").

Example 24.27. Assume that k is algebraically closed, and that A/k is a finite commutative k-algebra. We claim that the following are equivalent:

(i) A is separable over k.

(ii) A is reduced, i.e., $\mathrm{rad}(A) = 0$.
(iii) A is a product of copies of k.

Indeed, the equivalence of the first two conditions follows from Proposition 24.22(ii), and the equivalence of the second and the third conditions follows from (102).

Proposition 24.28. *Given any finite commutative algebra A over a field F, and any field extension F/k (not necessarily algebraic), A is separable over k if and only if $A \otimes_k F$ is separable over F.*

Proof. This is immediate from the equality $[A : k]_s = [F \otimes_k A : F]_s$ (Proposition 24.22(i)). □

Proposition 24.29. *For a finite commutative k-algebra A and an algebraically closed field L containing k, the following are equivalent and independent of L:*

(i) A is separable over k, i.e., $\#\,\mathrm{Hom}_{k\text{-}Alg}(A, L) = [A : k]$.
(ii) For some or equivalently any algebraically closed field L containing k, $A \otimes_k L$ is reduced, i.e., semisimple; equivalently, A is 'absolutely semisimple' or 'geometrically semisimple', or 'geometrically reduced.
(iii) $A \otimes_k L$ is a product of copies of L.
(iv) $A \otimes_k K$ is reduced for any field extension K/k.
(v) The symmetric bilinear form $\mathrm{tr} = \mathrm{tr}_{A/k} : A \times A \to k$ is nondegenerate.
(vi) $\mathrm{Hom}_{k\text{-}Alg}(A, L) \subset \mathrm{Hom}_k(A, L)$ is an L-vector space basis for $\mathrm{Hom}_k(A, L)$.

Slogan: Separable = 'absolutely semisimple' = 'geometrically semisimple' = 'geometrically reduced'.

Exercise 24.30. Example/exercise: Assuming Proposition 24.29, show that if V is a finite dimensional vector space over k, and $T \in \mathrm{End}_k(V)$, then T is semisimple over V – i.e., diagonalizable over some algebraic closure of k – if and only if the k-algebra $k[T] \subset \mathrm{End}_k(V)$ is separable.

Remark 24.31. It seems instructive to note that each of the conditions in Proposition 24.29 is invariant under replacing A and k by $A \otimes_k L$ and L. [69] This claim is enough to see for the conditions (i), (iv), (v) and (vi). In the case of (i), this follows from Proposition 24.28. In the case of (iv), this follows from the fact that any field is contained in an algebraically closed field, and tensoring with a field extension does not kill any element, and in particular preserves nilpotents. In the case of (v), this follows from the fact that the nondegeneracy of a bilinear form can be detected by its determinant, which may be computed after a base-change. For the condition (vi), this follows from the identities $\mathrm{Hom}_{k\text{-}Alg}(A, L) \cong \mathrm{Hom}_{k\text{-}Alg}(A \otimes_k L, L)$ and $\mathrm{Hom}_k(A, L) \cong \mathrm{Hom}_L(A \otimes_k L, L)$, which follow from Hom-tensor adjointness.

[69] And indeed, the proof will show how things simplify when we can base-change to an algebraically closed field. Again, this illustrates the flexibility involved in considering algebras, where we can take tensor products.

Proof of Proposition 24.29. Once we prove the equivalence of the conditions, it will follow that they are independent of L, since (i) is. Further, by Remark 24.31, we may replace A with $A \otimes_k L$ and assume without loss of generality that $k = L$ is algebraically closed (though this assumption will not be used when we deal with (vi)).

The equivalence of the first three conditions follows from Example 24.27. Since (i) is independent of L, it follows that so is (ii). Since every field is contained in an algebraically closed field, it is now easy to see that (ii) is equivalent to (iv). I am tempted to call this as "separable = universally reduced", but that is probably bad terminology.

Now let us prove the equivalence of these conditions with (v). If $A = \prod_{i=1}^r A_i$ with each A_i local, it is immediate that the bilinear form $\mathrm{tr}_{A/k}$ is the orthogonal sum of the $\mathrm{tr}_{A_i/k}$, and that A/k is separable if and only if each A_i/k is (because both $[A : k]$ and $[A : k]_s$ respect product decompositions in A). So we may and do assume that A is Artin local over $k = L$.

In this case, if A is separable, then $A = L$ (the equivalence of (i) and (iii)), and $\mathrm{tr}_{A/L} = \mathrm{tr}_{A/k}$ simply takes (a, b) to ab, which is clearly nondegenerate. On the other hand, if A is not separable, we know from the equivalence of (i) and (ii) that A contains a nonzero nilpotent element $a \in A$, which is clearly in the radical of $\mathrm{tr}_{A/k}$: for all $b \in B$, $ab \in A$ is nilpotent, so $m_{ab} : A \to A$, being a nilpotent linear operator, has trace zero.

The equivalence between (i) and (vi) follows because $\mathrm{Hom}_k(A, L)$ is an L-vector space of dimension $[A : k]$, and the elements of $\mathrm{Hom}_{k\text{-}Alg}(A, L)$ are linearly independent by Dedekind's linear independence of characters, Theorem 24.32 below, and hence span a subspace of dimension $[A : k]_s$. □

Theorem 24.32. *Let G be a monoid, and K a field. Let $\chi_1, \ldots, \chi_r$ be distinct characters (i.e., monoid homomorphisms, which by definition are required to take 1 to 1)* [70] *$G \to K$. Then $\chi_1, \ldots, \chi_r$ are linearly independent elements of the K-vector space* $\mathrm{Maps}(G, K)$ *of maps $G \to K$.*

Proof. If $r = 1$, there is nothing to prove, since $\chi_1 \neq 0$ (since it sends $1 \in G$ to $1 \in K$).

Now suppose $r > 1$ and that $\chi_1, \ldots, \chi_r$ is a minimal collection of linearly dependent characters, and say

$$\sum_{i=1}^{r} a_i \chi_i = 0 \tag{104}$$

for some $a_1, \ldots, a_r \in K$, not all zero – in fact, each nonzero by minimality. Since $\chi_1 \neq \chi_2$, let $z \in G$ be such that $\chi_1(z) \neq \chi_2(z)$. Then, since the above equation remains true after replacing each χ_i by $(g \mapsto \chi_i(zg)) = \chi_i(z)\chi$, we get

$$\sum_{i=1}^{r} a_i \chi_i(z) \chi_i = 0. \tag{105}$$

[70] In fact, it seems that we can do with much weaker assumptions: G just needs to have some binary operation, and the χ_i just need to each intertwine the operation on G with multiplication in K, each be nonzero, and be pairwise distinct.

Multiplying (104) by $\chi_1(z)$ and subtracting (105), we get

$$\sum_{i=2}^{r} a_i(\chi_i(z) - \chi_1(z))\chi_i = 0,$$

which is a nontrivial relation as $\chi_2(z) \neq \chi_1(z)$ and a_2 is nonzero. This contradicts the minimality of r. □

Corollary 24.33. *(i) Subalgebras, quotients (i.e., by an ideal), products and tensor products of finite separable algebras over k are separable (thus, being a quotient, the image of any finite separable k-algebra under a k-algebra homomorphism is separable too).*

(ii) If $\{A_i\}_i$ is a collection of separable algebras contained in a finite commutative k-algebra A, then the subalgebra of A generated by the A_i's is separable.

(iii) If A is a finite commutative algebra over k, then there is a k-subalgebra A_s of A such that A_s is separable over k, and such that any other separable k-subalgebra B of A is contained in A_s.

(iv) An algebraic field extension K/k is separable if and only if K is generated over k generated by a family of elements that are separable over k, and equivalently if and only if it is generated by a family of subextensions that are separable over k.

(v) If K/k is an algebraic field extension, $E, F \subset K$ are subextensions of K/k, and E/k is separable, then so is EF/F.

(vi) If $K/E/k$ is a chain of field extensions with K/k algebraic, then K/k is separable if and only if K/E and E/k are.

Remark 24.34. Once we prove (i) of the proposition, it follows that a commutative finite k-algebra is separable if and only if each finite k-subalgebra of it is, so that the definition of separability for infinite field extensions given in Definition 24.24 can also be applied to finite field extensions. Please keep this in mind while reading the following proof.

Proof of Corollary 24.33. In (i), the assertion about subalgebras and products follows from the characterization of separable algebras (among commutative finite k-algebras) as the ones that remain reduced on tensoring with an algebraically closed field (the condition in (ii) of Proposition 24.29). The assertion about quotients and tensor products is easy to see from the condition (iii) of the same proposition (use that a quotient of a product of copies of L is a product of copies of L, and that taking products of algebras commutes with taking their tensor product with a fixed algebra). This proves (i).

For (ii), the subalgebra of A generated by the A_i's is, by finite dimensionality, generated by some finite subcollection $A_{i_1}, \ldots, A_{i_n}$ of them, and is hence the image of the obvious homomorphism $A_{i_1} \otimes_k \otimes \cdots \otimes_k A_{i_n} \to A$. Therefore, its separability follows from (i).

For (iii), consider the subalgebra of A generated by all its separable subalgebras, and apply (ii).

Now we come to (iv). Suppose K/k is an algebraic field extension that is generated by a family $\{K_i/k\}_i$ of separable subextensions. Without loss of generality, each K_i/k is a

finite separable extension. We need to see that each finite subextension K'/k of K/k is separable, which follows because K' is contained in the subfield of K generated by finitely many of finite subextensions of the K_i's (and then (ii) applies). On the other hand, if K is generated over k by elements α that are separable over k, then K is generated by the various $k[\alpha]/k$, each of which is separable. This completes the proof of (iv).

Now let us prove (v). When E/k is finite, $E \otimes_k F$ is separable over F by Proposition 24.22(i), and then the separability of EF/F follows from applying the assertion about images in (i) to the multiplication map $E \otimes_k F \to EF$ of F-algebras. The general case then easily follows using (iv).

When K/k is finite, (vi) follows from the multiplicativity of the separable degree (Proposition 24.18). We leave the general case as an exercise – there is a little bit of work to do (or see Theorem 4.5 of Chapter V of Serge Lang's book). □

Exercise 24.35. Let $A = \prod_{i=1}^r A_i$ be a finite commutative algebra over k, with each A_i local. Let $K_i = A_i/\operatorname{rad}(A_i)$ for each i, so we have an isomorphism $A/\operatorname{rad}(A) \to \prod_{i=1}^r K_i$. Show that A is separable over k if and only if for each i, A_i equals K_i and is a separable field extension of k. Thus, finite separable algebras are the same as finite products of separable extensions of k.

24.4. **Some more characterizations of separable algebras.** The following proposition is pilfered from Qiaochu Yuan's blog (link given below), except that we have added in our restrictive assumptions regarding commutativity and finiteness.

Proposition 24.36. *If A/k is a finite commutative k-algebra, then the following are equivalent:*

(i) A is separable.
(ii) $A \otimes_k A$ is reduced.
(iii) A is projective as an (A, A)-bimodule (i.e., as a module over $A \otimes_k A^{op} = A \otimes_k A$).
(iv) There is a decomposition $A \otimes_k A \cong A \times A'$ of rings, under which (viewed as an identification) $m : A \otimes_k A \to A$ becomes the projection $A \times A' \to A'$ (the idempotent implicated in this decomposition, associated to A, is called the separability idempotent).

Remark 24.37. (ii) of the proposition implies that, to check if a finite extension K/k is separable, one does not need to compute $K \otimes_k L$ for an algebraically closed field L containing k; just computing $K \otimes_k K$ is enough.

Proof of Proposition 24.36. If A is separable, then $A \otimes_k A$ is separable by Corollary 24.33, and in particular reduced. Thus, (i) implies (ii).

If (ii) is satisfied, then $A \otimes_k A$, being Artinian and reduced, is a product of fields and hence semisimple. Since every module over a semisimple ring is projective, A is projective as an (A, A)-bimodule, giving (iii). Thus, we have proved (ii) ⇒ (iii).

Now assume (iii). Since $m : A \otimes_k A \to A$ is surjective, the projectivity of A lets us write $A \otimes_k A$ as $I \oplus I'$, where m restricts to an isomorphism $I \to A$, and I' is a complement. But then I and I' are rings in their own right with multiplication inherited from $A \otimes_k A$, so calling these rings A and A', we get the decomposition $A \otimes_k A = A \times A'$. Thus, (iii) $\Rightarrow$ (iv).

Note that this argument can be reversed, so the projectivity of A is equivalent to m having a section; so in fact we have (iii) $\iff$ (iv).

Finally, let us prove (iv) $\Rightarrow$ (i). If A is projective as an (A, A)-bimodule, or equivalently as an $A \otimes_k A$-module (recall that $A = A^{op}$ by commutativity), then $A_L := A \otimes_k L$ is projective as an $A_L \otimes_L A_L$ bimodule. Thus, to prove that A is separable we reduce to the case where $k = L$ is algebraically closed. It is easy to further reduce this to the case where A is Artin local. But since $k = L$ is algebraically closed, it is easy to check that if A is Artin local, then so is $A \otimes_k A$, so if $m : A \otimes_k A \to A$ induces a decomposition $A \otimes_k A \cong A \times A'$, then $m : A \otimes_k A \to A$ is an isomorphism. Comparing dimensions, we conclude that $A = k$, and hence A is reduced over the algebraically closed field $L = k$, and hence separable. This proves (iv) $\Rightarrow$ (i), as desired. $\square$

Remark 24.38. Later, if we get time, we might in a future lecture relate the conditions above involving $m : A \otimes_k A \to A$ to a different criterion for separability, in terms of Kähler differentials vanishing. But it looks like we may not get time for that.

The above proposition applies even when k is a commutative ring and A is a possibly non-commutative k-algebra. See https://qchu.wordpress.com/2016/03/27/separable-algebras/

25. Lecture 25 – Finite Galois theory, classical proof

In this lecture, we will define Galois extensions, and prove a classical form of the Galois correspondence (for finite Galois groups), (proving and) using the primitive element theorem. This is different from Artin's proof in his Notredame lectures, which will be taken up in Lecture 26 – I find that proof more enlightening, but I hadn't absorbed and framed it to my taste when I actually gave Lecture 25.

25.1. Separable closure, the category of finite separable k-algebras.

Definition 25.1. (i) A field k is called separably closed if it has no finite separable field extension K/k.

(ii) A field extension $k \hookrightarrow K$ is said to be a separable closure of k if K/k is algebraic and K is separably closed.

Remark 25.2. It is clear that every algebraically closed field is separably closed, but the converse is not true. Indeed, if $k = F(t)$, where F is any field of charcteristic p, then $x^p - t \in k[x]$ does not have a root in a separable closure of k, so a separable closure of k cannot be algebraically closed.

Lemma 25.3. *(i) Every field k has a separable closure. In fact, if $k \hookrightarrow L$ is an algebraic closure, and $K = \{\alpha \in L \mid \alpha \text{ is separable over } k\}$, then K is a separable closure of k.*

(ii) If $k \hookrightarrow K$ is a separable closure, every separable polynomial in $k[x]$ splits completely in K. Moreover, K is a splitting field of the family of all irreducible separable polynomials in $k[x]$.

(iii) If $k \hookrightarrow K$ is a separable closure, $K \hookrightarrow L$ any field extension, and E/k is a separable algebraic field extension, then every k-algebra homomorphism $E \to L$ has image in K. We have $[E : k]_s = \#\operatorname{Hom}_{k\text{-}Alg}(E, k^s)$.

(iv) Conversely to (i), given a separable closure K/k, and any algebraic closure $K \hookrightarrow L$ of K, the composite $k \hookrightarrow K \hookrightarrow L$ is also an algebraic closure, and we have $K = \{\alpha \in L \mid \alpha \text{ is separable over } k\}$.

(v) If $k \hookrightarrow K$ is a separable closure and E/k is a separable algebraic extension, then there exists a k-algebra embedding $E \hookrightarrow K$.

(vi) A separable closure of k is unique up to a nonunique isomorphism.

Proof. Let us prove (i) (so assume its setting). Clearly K/k is algebraic. If K were not separably closed, it would have a nontrivial separable (algebraic) extension $K \hookrightarrow K_1$, and then since L is algebraically closed we would get a K-algebra homomorphism $\sigma : K_1 \hookrightarrow L$, with $\sigma(K_1) \supsetneq K$. Then, given any $\alpha \in \sigma(K_1) \backslash K$, since $K[\alpha]/K$ and K/k are separable, we would get that $K[\alpha]/k$ is separable (the tower property, Corollary 24.33(vi) from Lecture 24), so that $\alpha \in L$ would be separable over k, so $\alpha \in K$, a contradiction. This proves (i).

(ii) is an easy consequence of (i) and (iii) is an easy consequence of (ii), so we come to (iv). Clearly L/k is algebraic, and hence $k \hookrightarrow L$ is an algebraic closure. By (i),

$K' := \{\alpha \in L \mid \alpha \text{ is separable over } k\}$ is a separable closure of k, which contains K. But this implies that K'/K is separable. Since K is separably closed, we conclude that $K = K'$.

Let us prove (v). Choosing an algebraic closure $K \hookrightarrow L$ as in (iv), we know that there is a k-algebra embedding $E \hookrightarrow L$, which has image in K by (iii). This proves (v), and (vi) is immediate from (v). □

Notation 25.4. (i) In what follows, given a field k, when we write k^s, it will be understood that a separable closure $k \hookrightarrow k^s$ has been chosen, with k^s as its underlying field.

(ii) We will write $\mathrm{Gal}(k^s/k)$ for the group $\mathrm{Aut}_{k\text{-}Alg}(k^s)$ of k-algebra automorphisms of k^s.

(iii) Let $f\acute{e}t_k$ denote the category of finite commutative k-algebras which are separable over k (of course, with morphisms being k-algebra homomorphisms). To explain the notation, a finite separable k-algebra is the same as what is called a finite étale k-algebra.

Already we can see an advantage that the separable closure has, which the algebraic closure does not:

Corollary 25.5. *The inclusion $k \hookrightarrow (k^s)^{\mathrm{Gal}(k^s/k)}$ is an isomorphism.*

Proof. If on the contrary $[(k^s)^{\mathrm{Gal}(k^s/k)} : k] = [(k^s)^{\mathrm{Gal}(k^s/k)} : k]_s > 1$, then by the equality $[(k^s)^{\mathrm{Gal}(k^s/k)} : k]_s = \#\,\mathrm{Hom}_{k\text{-}Alg}((k^s)^{\mathrm{Gal}(k^s/k)}, k^s)$ (Lemma 25.3(iii)), there exists a k-algebra embedding $(k^s)^{\mathrm{Gal}(k^s/k)} \hookrightarrow k^s$ that is different from the inclusion. Therefore, by Lemma 25.3(v), it extends to a k-algebra embedding $k^s \to k^s$, which is automatically an isomorphism (see the first exercise of Lecture 24), and is non-identity. In other words, we get an element of $\mathrm{Gal}(k^s/k)$ that is not the identity on $(k^s)^{\mathrm{Gal}(k^s/k)}$, a contradiction. □

25.2. The category of finite split k-algebras.

Definition 25.6. By a finite split k-algebra, we mean a finite product of copies of k, each such product $k \times \cdots \times k$ being viewed as a k-algebra by the 'diagonal' embedding. Let $(spl(k))_f$ denote the category of finite split k-algebras (and k-algebra homomorphisms between them).

Lemma 25.7. *If k is separably closed, then a finite commutative k-algebra A is separable over k if and only if A is finite split. Thus, when k is separably closed, the inclusion of the full subcategory $(spl(k))_f$ into $f\acute{e}t_k$ is an equivalence (even an isomorphism).*

Proof. We saw in Lecture 24 (Exercise 24.35) that a finite commutative k-algebra A is separable over k if and only if A is a finite product $\prod_{i=1}^r K_i$ of finite separable field extensions K_i of k. When k is separably closed, this is by definition equivalent to A being a finite product of copies of k. □

Remark 25.8. Lemma 25.7 gives a different, and perhaps better, proof of Lemma 25.3(v) when E/k is finite. Namely, if E/k is finitely generated, then $E \otimes_k K$, being separable over K (see Proposition 24.28 from Lecture 24), is a product of copies of K by Lemma 25.7. This implies that there exists a ring homomorphism $E \otimes_k K \to K$, and hence a ring homomorphism $E \hookrightarrow E \otimes_k K \to K$, which is the desired embedding. The general case can be deduced using a colimit argument, but we skip the details. Probably some other assertions of Lemma 25.3 can also be proved in analogous ways, but I have not thought about it.

Exercise 25.9. (i) (Simple, but important). Show that $spl(k)_f^{op}$ is equivalent to the category $FinSet$ of finite sets: in fact, we have a functor $FinSet \rightsquigarrow spl(k)^{op}$ given by $X \mapsto \mathrm{Maps}(X, k)$, and a functor $spl(k)_f^{op} \rightsquigarrow FinSet$ given by $A \mapsto X_A := \mathrm{Hom}_{k\text{-}Alg}(A, k)$; show that these are mutually quasi-inverse equivalences of categories.

(ii) Conclude from your proof of (i) that "any finite split k-algebra is canonically split": if $A \in \mathrm{Ob}\, spl(k)_f^{op}$, then we have an isomorphism of k-algebras

$$Gelf := \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A,k)} \sigma : A \to \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A,k)} k = \mathrm{Maps}(X_A, k), \tag{106}$$

where $X_A := \mathrm{Hom}_{k\text{-}Alg}(A, k)$.

Here, the notation '$Gelf$' is used because we may think of (106) as a Gelfand transform: it is obtained from the tautological evaluation pairing:

$$A \times \mathrm{Hom}_{k\text{-}Alg}(A, k) \to k,$$

and is a very trivial analogue of the Gelfand transform for commutative C^*-algebras, where the analogue of $X_A = \mathrm{Hom}_{k\text{-}Alg}(A, k)$ is the compact Hausdorff space of nonzero $*$-homomorphisms from A to $\mathbb{C}$.

Note that for more general finite commutative k-algebras A, the map of (106) will not be called Gelfand transform, and we will not denote it by $Gelf$.

25.3. **Another property of finite split k-algebras.**

Proposition 25.10. *Let A be a finite commutative k-algebra. Then*

(i) A/k is separable (or equivalently $A \in \mathrm{Ob}\, f\acute{e}t_k$) if and only if $A \otimes_k k^s$ is a product of copies of k^s.

(ii) When the equivalent conditions of (i) hold, $A \otimes_k k^s$ is canonically a product of finite copies of k^s, via the isomorphism:

$$Gelf : A \otimes_k k^s \xrightarrow{\cong} \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A,k^s)} k^s, \tag{107}$$

whose σ-component is the map $A \otimes_k k^s \to k^s$ that sends $a \otimes b$ to $\sigma(a)b$, for each $a \in A$ and $b \in k^s$. [71]

[71] One can't of course, technically, define it this way: rather, one either uses the universal property of the tensor product to define it from the k-bilinear pairing $A \times k^s \to k^s$ given by $(a, b) \mapsto \sigma(a)b$, and

Proof. For (i), use that A is separable over k if and only if $A \otimes_k k^s$ is separable over k^s (use Proposition 24.28 from Lecture 24), and apply Lemma 25.7. Given this, (ii) follows from Exercise 25.9(ii), and the identification $\mathrm{Hom}_{k\text{-}Alg}(A, k^s) \to \mathrm{Hom}_{k^s\text{-}Alg}(A \otimes_k k^s, k^s)$ given by Hom-tensor adjointness (see Exercise 24.21 from Lecture 24). □

Notation 25.11. (i) We make $\mathrm{Gal}(k^s/k)$ act on $A \otimes_k k^s$ and $X_A := \mathrm{Hom}_{k\text{-}Alg}(A, k^s)$ in the obvious fashion (i.e., using its action on k^s).

(ii) Further, we will view $\mathrm{Maps}(X_A, k^s)$ as a k-algebra under pointwise addition and multiplication, and make $\mathrm{Gal}(k^s/k)$ act on it by $\sigma \cdot f(x) = \sigma(f(\sigma^{-1}(x)))$ – this is the obvious way, given a group G acting on sets X and Y, to make G act on $\mathrm{Maps}(X, Y)$. Note that (107) can be written:

$$Gelf : A \otimes_k k^s \xrightarrow{\cong} \mathrm{Maps}(X_A, k^s), \tag{108}$$

and verify that (108) is $\mathrm{Gal}(k^s/k)$-equivariant – we will use this in what follows.

Proposition 25.12. *Let A/k be a finite separable k-algebra. Consider $X_A = \mathrm{Hom}_{k\text{-}Alg}(A, k^s)$ with its obvious $\mathrm{Gal}(k^s/k)$-action. Let $\mathrm{Maps}(X_A, k^s)^{\mathrm{Gal}(k^s/k)} = \mathrm{Maps}_{\mathrm{Gal}(k^s/k)}(X_A, k^s)$ be the algebra of $\mathrm{Gal}(k^s/k)$-equivariant maps $X_A \to k^s$, with pointwise addition and multiplication. Then we have the following isomorphism of k-algebras generalizing* (106)*:*

$$Gelf : A \to \mathrm{Maps}_{\mathrm{Gal}(k^s/k)}(X_A, k^s), \tag{109}$$

given by

$$A \ni a \mapsto (f \mapsto f(a)).$$

Note how this isomorphism is obtained from the evaluation pairing

$$A \times X_A = A \times \mathrm{Hom}_{k\text{-}Alg}(A, k^s) \to k^s.$$

Proof. Proposition 25.10 gave us an isomorphism of k^s-algebras

$$A \otimes_k k^s \to \mathrm{Maps}(X_A, k^s).$$

We have noticed that this map respects the actions of $\mathrm{Gal}(k^s/k)$ defined in Notation 25.11. Therefore, it restricts to an isomorphism

$$A = (A \otimes_k k^s)^{\mathrm{Gal}(k^s/k)} \to \mathrm{Maps}_{\mathrm{Gal}(k^s/k)}(X_A, k^s),$$

where the first equality is an immediate consequence of Corollary 25.5. □

25.4. **Galois correspondence – informal motivation.** In Exercise 25.9, we saw that finite split k-algebras have a nice description, as equivalent to the category of finite sets. In Galois theory, one sort of extends that to k-algebras that are finite separable, but not necessarily split (rather, they are just "k^s-split"). Please compare the constructions below to their analogues in Exercise 25.9.

---

checks that it is a morphism of algebras, or uses that the tensor product is a coproduct in the category of commutative k-algebras, to define it from the k-algebra homomorphisms $\sigma : A \to k^s$ and $\mathrm{id} : k^s \to k^s$.

Notation 25.13. (i) We will write $FinSet$ for the category of finite sets (and set-theoretic maps between them). If G is a group, G-$FinSet$ will denote the category of finite sets with a G-action (and G-equivariant morphisms between them).

(ii) We have a functor

$$\mathcal{F} : (f\acute{e}t_k)^{op} \rightsquigarrow \mathrm{Gal}(k^s/k)\text{-}FinSet, \quad A \rightsquigarrow X_A = \mathrm{Hom}_{k\text{-}Alg}(A, k^s) \tag{110}$$

(where X_A is viewed together with the $\mathrm{Gal}(k^s/k)$-action on it from Notation 25.11). Note that the 'op' is necessary to get the arrows in the correct direction.

(iii) In the opposite direction, we have a functor

$$\mathcal{G} : \mathrm{Gal}(k^s/k)\text{-}FinSet \to (f\acute{e}t_k)^{op}, \quad X \rightsquigarrow A_X := \mathrm{Maps}_{\mathrm{Gal}(k^s/k)}(X, k^s). \tag{111}$$

Note that since A_X is contained in a finite product of copies of k^s (namely, in $(k^s)^{\#X}$), it is separable over k. There is in fact some work to be done to show that A_X is indeed finite over k, but we will omit it since this subsection is informal motivation. [72].

(iv) There are obvious natural transformations $\eta : \mathrm{id}_{(f\acute{e}t_k)^{op}} \to \mathcal{G}\circ\mathcal{F}$ and $\epsilon : \mathrm{id}_{\mathrm{Gal}(k^s/k)\text{-}FinSet} \to \mathcal{G}\circ\mathcal{F}$: if A is a finitely generated k-algebra, then:

- $\eta_A : A \to \mathcal{G}(\mathcal{F}(A)) = \mathrm{Maps}_{\mathrm{Gal}(k^s/k)}(X_A, k^s)$ is given by $a \mapsto (f \mapsto f(a))$, and is hence exactly the isomorphism of Proposition 25.12.
- $\epsilon_X : X \to \mathcal{F}(\mathcal{G}(X)) = \mathrm{Hom}_{k\text{-}Alg}(A_X, k^s) = \mathrm{Hom}_{k\text{-}Alg}(\mathrm{Maps}_{\mathrm{Gal}(k^s/k)}(X, k^s), k^s)$ is given by $x \mapsto (a \mapsto a(x))$.

Thus, η_A and ϵ_A are both obtained from the obvious evaluation pairings.

However, unfortunately, unlike with Exercise 25.9, the functor $\mathcal{F}$ of Notation 25.4(ii) is not an equivalence of categories, since it (can be shown that it) is not essentially surjective. While η is a natural isomorphism by Proposition 25.12, ϵ is not. $\mathcal{F}$ can be shown to be fully faithful, and this together with the fact that $\eta : \mathrm{id}_{k\text{-}Alg} \rightsquigarrow \mathcal{G}\circ\mathcal{F}$ is a natural isomorphism is arguably the "easy half" of Galois theory. To prove the other half, or even to motivate its formulation, we will (probably) need to work with finite Galois extensions, a notion we will soon define.

The following is an informal statement of the Galois correspondence in this setting.

Theorem 25.14. *There is a topology on* $\mathrm{Gal}(k^s/k)$ *such that the functors* $\mathcal{F}$ *and* $\mathcal{G}$ *from* (110) *and* (111) *of Notation 25.13 define mutually quasi-inverse equivalences of categories*

$$(f\acute{e}t_k)^{op} \rightsquigarrow (\mathrm{Gal}(k^s/k)\text{-}FinSet)_{cts},$$

where the right hand side is the category of finite sets with a continuous $\mathrm{Gal}(k^s/k)$*-action.*

[72]Here is a sketch of how you can show it using material we will see later in this lecture. We immediately reduce to the case where $X = \mathrm{Gal}(k^s/k)/H$ for some finite index subgroup $H \subset \mathrm{Gal}(k^s/k)$. It is enough to show that $(k^s)^H$ is finite over k. Replacing H by a finite index subgroup that is normal in $\mathrm{Gal}(k^s/k)$, we may and do assume that $H \subset \mathrm{Gal}(k^s/k)$ is normal. Then $((k^s)^H)^{\mathrm{Gal}(k^s/k)/H} = k$ by Corollary 25.5, and Lemma 25.30 below shows that $(k^s)^H$ is a finite extension of k

We will come discuss the topology alluded to in the above theorem, Krull topology, only in Lecture 26. Meanwhile, let us informally discuss some of the motivation for it.

Remark 25.15. (i) One 'analogy-based' motivation-of-sorts for considering the functor $\mathcal{F}$ of (110) can be given as follows. Recall that if $\mathcal{A}$ is an abelian category, and if $\mathcal{A}$ satisfies certain conditions including the existence of a projective generator P, we had asserted without proof that $\mathcal{A}$ can be realized as a module category (see the discussion on Morita equivalence in Lecture 19): sending $A \in \mathrm{Ob}\,\mathcal{A}$ to $\mathrm{Hom}_{\mathcal{A}}(P, A)$, viewed as a module over $R := \mathrm{End}_{\mathcal{A}}(P)^{op}$, defines an equivalence of categories $\mathcal{A} \rightsquigarrow R\text{-}Mod$. Clearly, (110) is an analogue, where we have a G-set instead of an R-module because the category involved is not abelian. Thus, just as one might wish to study various abelian categories as module categories, it could be simplifying to study the seemingly lawless category $f\acute{e}t_k$ as a category of G-sets. However, there are some differences even excluding abelianness: e.g., k^s is not an object of $f\acute{e}t_k$, since k^s/k is not finite (rather, it is an 'Ind-object' of the category, for those who are interested).

(ii) Another way to motivate this is that on $spl(k)_f^{op}$, the functor $\mathcal{F}$ of (110) restricts to the equivalence of categories $spl(f)_k^{op} \rightsquigarrow FinSet$ in Exercise 25.9(i), and on the subcategory of finite $\mathrm{Gal}(k^s/k)$-sets consisting of sets with trivial action, the functor $\mathcal{G}$ of (111) restricts to the quasi-inverse equivalence of categories $FinSet \rightsquigarrow spl(f)_k^{op}$ from Exercise 25.9(i).

(iii) Let us expand on the above point. Recall that if $A = k[x_1, \ldots, x_n]/(g_1, \ldots, g_m)$ is a finitely generated k-algebra, then A can be thought of as a system of equations, whose solutions with entries in a k-algebra R are given by:

$$X(R) := \{(a_1, \ldots, a_n) \in R^n \mid g_i(a_1, \ldots, a_n) = 0 \,\forall\, 1 \leqslant i \leqslant m\}.$$

Here, we may think of $R \rightsquigarrow X(R)$, viewed as a functor $X : k\text{-}Alg \rightsquigarrow Set$, as the 'variety of solutions to the system of equations A'. Thus, the 'solutions' functor associated to a k-algebra A is just $\mathrm{Hom}_{k\text{-}Alg}(A, -)$. Therefore, Exercise 25.9 says that a split k-algebra is completely determined by the value of the solutions functor at the k-algebra k. This is not true if A properly contains k: e.g., $\mathrm{Hom}_{k\text{-}Alg}(A, k) = \varnothing$ if A is a finite separable field extension of k. In Proposition 25.10, we observe that when A/k is finite separable, $A \otimes_k k^s$ is entirely captured by $\mathrm{Hom}_{k\text{-}Alg}(A, k^s)$, which is the solution functor evaluated at the k-algebra k^s. If we want to capture A itself, and not just $A \otimes_k k^s$, the point is that we need to take care of the $\mathrm{Gal}(k^s/k)$-action on $A \otimes_k k^s$, and recover A as the set of $\mathrm{Gal}(k^s/k)$-fixed points in $A \otimes_k k^s$: in other words, we are still looking at the solution functor $X(k^s) := \mathrm{Hom}_{k\text{-}Alg}(A, k^s)$ evaluated at k^s, but in the process also keeping track of the action on $X(k^s)$ by $\mathrm{Aut}_{k\text{-}Alg}(k^s) = \mathrm{Gal}(k^s/k)$.

(iv) Then, of course, there is the mandatory analogy with covering spaces. Since we are considering $(f\acute{e}t)_k^{op}$, a k-algebra A would be viewed as a map $\mathrm{Spec}\, A \to \mathrm{Spec}\, k$, where $\mathrm{Spec}\, A$ and $\mathrm{Spec}\, k$ are just A and k, viewed as objects in the opposite category. The isomorphism $A \otimes_k k^s \to \prod_\sigma k^s$ from Proposition 25.10 could be thought of

as saying that $\operatorname{Spec} A \to \operatorname{Spec} k$ is a 'covering map', that becomes a trivial covering when pulled back to the 'universal cover $\operatorname{Spec} k^s \to \operatorname{Spec} k$'.

Thus, the equivalence of categories in Theorem 25.14 is analogous to how the category of coverings of a topological space X is equivalent to the category of finite sets with a $\pi_1(X, x)$-action, where $\pi_1(X, x)$ is the fundamental group of X at some base point x.

In fact, there is a common description that can capture both these theories, involving the notion of 'Galois categories' of Grothendieck. I planned to discuss it in Lecture 26, but could not. If I get time I might write something about it in the notes for Lecture 26.

25.5. **Normal extensions.**

Theorem 25.16. *Let K/k be an algebraic (but not necessarily finite) extension of k, contained in an algebraic closure L of k. The following are equivalent:*

(NOR 1) $\sigma(K) = K$ *for all* $\sigma \in \operatorname{Hom}_{k\text{-}Alg}(K, L)$.
(NOR 2) K *is a splitting field of a family of polynomials in* $k[x]$.
(NOR 3) Every irreducible polynomial in $k[x]$ *which has a root in* K*, splits into linear factors in* $K[x]$.

Proof. (NOR3) $\Rightarrow$ (NOR 2) is immediate, since one can take a family of polynomials as in (NOR 2) to be the family of all irreducible polynomials in $k[x]$ with a root in K. (NOR2) $\Rightarrow$ (NOR1) follows from the fact that a family of polynomials in $k[x]$ has a unique splitting field within a given algebraically closed field L, and that any $\sigma \in \operatorname{Hom}_{k\text{-}Alg}(K, L)$ clearly necessarily takes a splitting field of a family of polynomials to a splitting field of the same family.

Now let us prove (NOR1) $\Rightarrow$ (NOR3). If $f \in k[x]$ is irreducible and has a root α in K, but does not split into linear factors in K, then f has a root $\alpha' \in L \backslash K$. We have an isomorphism of k-algebras $k[\alpha] \to k[\alpha'] \hookrightarrow L$ sending α to α', extending which to K we get a k-algebra embedding $\sigma : K \hookrightarrow L$ that sends α to α', and hence does not satisfy $\sigma(K) = K$. □

Definition 25.17. (i) An algebraic field extension K/k that satisfies the equivalent conditions of Theorem 25.16 is said to be a normal extension.

(ii) An algebraic extension K/k is said to be Galois if it is both normal and separable. If K/k is Galois, we will write $\operatorname{Gal}(K/k)$ for $\operatorname{Aut}_{k\text{-}Alg}(K)$ (this agrees with the notation $\operatorname{Gal}(k^s/k)$ defined earlier in Notation 25.4, as xample 25.18(ii) below shows).

(iii) If $f \in k[x]$ is a separable polynomial, so that any splitting field k_f of f is Galois over k, we define the Galois group of f over k to be $\operatorname{Gal}(k_f/k)$. Note that it is unique up to isomorphism.

Example 25.18. .

(i) Any quadratic extension K/k is normal, but it may or may not be separable. When $\operatorname{char} k \neq 2$, any quadratic extension K/k is immediately seen to be of the form

$K = k[\sqrt{\delta}]$ for some $\delta \in k$, and since irreducible polynomials of the form $x^2 - \delta$ are clearly separable when $\operatorname{char} k \neq 2$, it follows that any quadratic extension is separable and hence also Galois in this case.

Now assume that $\operatorname{char} k = 2$. Then a quadratic extension K/k may or may not be separable: if $K = k[\sqrt{\delta}]$ for some $\delta \in k$, K/k is not separable (as is the case for the extension $\mathbb{F}_2(t) \hookrightarrow \mathbb{F}_2(t)$ given by $t \mapsto t^2$), but if the quadratic extension K/k is obtained by adjoining a root of an irreducible polynomial $x^2 + ax + b$, then it is clearly separable and hence Galois as long as $a \neq 0$ (i.e., as long as the derivative of $x^2 + ax + b$ does not vanish).

(ii) If $k \hookrightarrow k^s$ is a separable closure and $k \hookrightarrow L$ is an algebraic closure, then k^s/k and L/k are normal. k^s/k is Galois, while L/k is Galois if and only if L/k is a separable closure. Thus, $\operatorname{Gal}(k^s/k)$ agrees with what we defined it to be in Notation 25.4.

Proposition 25.19. *(i) If $K/E/k$ are field extensions and K/k is normal, then K/E is normal.*

(ii) If E, F are fields containing a field k and contained in a field K, and E/k is a normal extension, then EF/F is a normal extension (here $EF \subset K$ is the compositum of E and F in K, i.e., the subfield of K generated by E and F).

(iii) If E, F are fields containing a field k and contained in a field K, and if E/k and F/k are normal extensions, so are EF/k and $(E \cap F)/k$.

Proof. Easy exercise. □

Lemma 25.20. *If K/k is a normal extension, $k \subset E \subset K$ is an intermediate subextension, and L is an algebraically closed field containing K, then we have maps (with the marked descriptions and properties):*

$$\begin{array}{ccc} \operatorname{Aut}_{k\text{-}Alg}(K) & \xrightarrow[\text{surjection}]{\text{restriction}} & \operatorname{Hom}_{k\text{-}Alg}(E, K) \\ {\scriptstyle (K\hookrightarrow L)\circ -}\downarrow{\scriptstyle \text{bijection}} & & {\scriptstyle \text{bijection}}\downarrow{\scriptstyle (K\hookrightarrow L)\circ -} \\ \operatorname{Hom}_{k\text{-}Alg}(K, L) & \xrightarrow[\text{surjection}]{\text{restriction}} & \operatorname{Hom}_{k\text{-}Alg}(E, L). \end{array}$$

Proof. The vertical arrows are clearly injections. That the left vertical arrow is a bijection follows from "(NOR 1)" in the definition of a normal extension. The bottom horizontal arrow is a surjection since L is algebraically closed.

Thus, $\operatorname{Aut}_{k\text{-}Alg}(K) \to \operatorname{Hom}_{k\text{-}Alg}(K, L) \to \operatorname{Hom}_{k\text{-}Alg}(E, L)$ is surjective, forcing the right vertical arrow to be surjective as well, and hence (being injective) bijective. Since the vertical arrows are bijective and the bottom horizontal arrow is surjective, the top horizontal arrow is surjective. □

Corollary 25.21. *In the setting of Lemma 25.20, if E/k is a normal extension as well, then $\operatorname{Aut}_{k\text{-}Alg}(K) \to \operatorname{Aut}_{k\text{-}Alg}(E)$ is surjective, with kernel $\operatorname{Aut}_{E\text{-}Alg}(K)$.*

Proof. Immediate. □

Example 25.22. However, unlike with separable extensions, normal extensions do not have the 'tower property': if $E/F/k$ are field extensions, and if E/F and F/k are normal, E/k may not be normal. For instance, $\mathbb{Q}[\sqrt[4]{2}]/\mathbb{Q}[\sqrt{2}]$ and $\mathbb{Q}[\sqrt{2}]/\mathbb{Q}$ are normal (being quadratic), but $\mathbb{Q}[\sqrt[4]{2}]/\mathbb{Q}$ is not normal.

Proposition 25.23. . *Let K/k be a Galois extension. For any subextension E/k of K/k, K/E is is Galois (so we can talk about* $\mathrm{Gal}(K/E)$*, as we will do below).*

Proof. K/E is separable since K/k is (the tower property for separable extensions), and K/E is normal by Proposition 25.19. □

Lemma 25.24. *For any finite field extension K/k, setting $G = \mathrm{Aut}_{k\text{-}Alg}(K/k)$, the following are equivalent:*

(i) K/k is Galois.
(ii) $\#G = [K : k]$.

Proof. Let L be an algebraically closed field containing K. We have

$$\#G = \#\operatorname{Aut}_{k\text{-}Alg}(K) \leqslant \#\operatorname{Hom}_{k\text{-}Alg}(K, L) = [K : k]_s \leqslant [K : k].$$

The first inequality is an equality if and only if K/k is normal, and the second inequality is an equality if and only if K/k is separable. Thus, the two conditions are equivalent. □

25.6. **Easy(?) half of finite Galois theory, classical version.** Here is the classical version of one (easier?) half of Galois theory:

Theorem 25.25. *(i) Let K/k be an arbitrary (not necessarily finite) Galois extension, and let $G = \mathrm{Gal}(K/k)$. Consider the map*

(112)
$$\Big\{\textit{Subextensions } E/k \textit{ of } K/k\Big\} \to \Big\{\textit{Subgroups of } \mathrm{Gal}(K/k)\Big\}, \quad E \mapsto H_E := \mathit{Fix}_G(E) = \mathrm{Gal}(K/E)$$

(where $Fix_G(E)$ is the subgroup of G that fixes the elements of E pointwise), and in the inverse direction,

(113)
$$H \mapsto K^H.$$

Then (112) *is injective (and in particular a bijection onto its image), and* (113) *is a left-inverse to it, i.e., for any subextension E/k of K/k, $K^{H_E} = E$ (more precisely, the obvious inclusion $E \hookrightarrow K^{H_E}$ is an isomorphism).*

(ii) Given a subextension E/k of K/k, E/k is Galois if and only if $\mathrm{Gal}(K/E) \subset \mathrm{Gal}(K/k)$ is a normal subgroup, in which case restriction induces a well-defined map $\mathrm{Gal}(K/k) \to \mathrm{Gal}(E/k)$, which quotients to an isomorphism $\mathrm{Gal}(K/k)/\mathrm{Gal}(K/E) \to \mathrm{Gal}(E/k)$.

Remark 25.26. If K/k is finite, then (112) is in fact bijective with (113) as a two-sided inverse: this is the other half of Galois theory for finite extensions, proved in Theorem 25.29 below.

The following lemma is a special case of the theorem, but also summarizes the non-formal/non-book-keeping 'Galois theory' input into the proof of the theorem:

Lemma 25.27. *If K/k is a Galois extension, then $k = K^{\mathrm{Gal}(K/k)}$.*

Proof. Let $K \hookrightarrow k^s$ be a separable closure. Since K/k is separable algebraic, $k \hookrightarrow K$ is a separable closure as well.

Since the restriction of any element of $\mathrm{Gal}(k^s/k)$ to K is an element of $\mathrm{Gal}(K/k)$ by the normality of K/k, we have

$$k \hookrightarrow K^{\mathrm{Gal}(K/k)} \hookrightarrow (k^s)^{\mathrm{Gal}(k^s/k)} = k$$

by Corollary 25.5 (and the fact that the $\mathrm{Gal}(k^s/k)$ in it has been observed to agree with the definition in Definition 25.17). Thus, the lemma follows. □

Proof of Theorem 25.25. For (i), it is enough to prove that the obvious inclusion $E \hookrightarrow K^{H_E}$ is an equality. But since $H_E = \mathrm{Gal}(K/E)$, this follows from Lemma 25.27.

Now we come to (ii). E/k is automatically separable by the tower property for separable extensions, and hence it is Galois if and only if it is normal.

Given $\sigma \in \mathrm{Gal}(K/k)$, since (112) is injective by (i), it follows that $\sigma(E) = E$ if and only if, inside $\mathrm{Gal}(K/k)$, the subgroups $\mathrm{Gal}(K/E)$ and $\mathrm{Gal}(K/\sigma(E))$, the latter of which is immediately verified to be $\sigma\,\mathrm{Gal}(K/E)\sigma^{-1}$, are equal. Varying σ over $\mathrm{Gal}(K/k)$, we see that $\mathrm{Gal}(K/E) \subset \mathrm{Gal}(K/k)$ is a normal subgroup if and only if $E \subset K$ is stabilized by $\mathrm{Gal}(K/k)$. It is an easy exercise to see using Lemma 25.20 that the latter condition is equivalent to E/k being normal.

Assuming that this condition holds, i.e., E/k is normal, the rest of (ii) follows from Corollary 25.21. □

25.7. **Primitive element theorem and the other half.** Right now, for the other half of finite Galois theory, we will use the primitive element theorem (and seems to be vaguely sort of equivalent to it). In Lecture 26 another (hopefully better) approach, involving Galois descent, will be discussed.

Theorem 25.28. *If K/k is a finite separable field extension, there exists $\alpha \in K$ such that $K = k[\alpha]$.*

Proof. If k is finite, so that K is finite as well, any generator α of the multiplicative group $K^\times$ will do: it is well-known that any finite subgroup of the multiplicative group of units of a field is cyclic. Thus, let us assume that k is infinite.

By induction, it suffices to show that if $\alpha, \beta \in K$, then $k(\alpha, \beta) \subset K$ is generated by a single element. Fix an algebraically closed field L containing K, and write $\sigma_1, \ldots, \sigma_n$ for the collection of k-algebra embeddings $k(\alpha, \beta) \to L$. Since $k(\alpha, \beta)/k$ is separable, we have $n = [k(\alpha, \beta) : k]$. It suffices to find $\gamma \in k(\alpha, \beta)$ such that $\sigma_i(\gamma) \neq \sigma_j(\gamma)$ for all $1 \leqslant i < j \leqslant n$: that will give $[k(\gamma) : k] \geqslant [k(\gamma)_s : k] \geqslant n = [k(\alpha, \beta) : k]$, so that $k(\gamma) = k(\alpha, \beta)$.

Consider the polynomial

$$P(x) := \prod_{1 \leqslant i < j \leqslant n} (\sigma_i(\alpha) + x\sigma_i(\beta) - \sigma_j(\alpha) - x\sigma_j(\beta)).$$

For all $1 \leqslant i < j \leqslant n$, we have either $\sigma_i(\alpha) \neq \sigma_j(\alpha)$ or $\sigma_i(\beta) \neq \sigma_j(\beta)$. From this, it is easy to see that P is not the zero polynomial. Since k is not finite, there exists $c \in k$ such that $P(c) \neq 0$, which translates to $\prod_{1 \leqslant i < j \leqslant n}(\sigma_i(\alpha + c\beta) - \sigma_j(\alpha + c\beta)) \neq 0$. Hence, we can simply take $\gamma = \alpha + c\beta$. □

Now we can prove the (remaining half of the) main theorem of Galois theory, in the case of finite extensions:

Theorem 25.29. *Let K/k be a finite Galois extension, and let $G = \mathrm{Gal}(K/k)$.*

(i) We have a bijection

(114)
$$\Big\{\textit{Subextensions } E/k \textit{ of } K/k\Big\} \to \Big\{\textit{Subgroups of } \mathrm{Gal}(K/k)\Big\}, \quad E \mapsto H_E := \mathit{Fix}_G(E) = \mathrm{Gal}(K/E),$$

with a two-sided inverse

(115)
$$H \mapsto K^H.$$

Thus, the obvious inclusions $E \hookrightarrow K^{H_E}$ and $H \hookrightarrow H_{K^H} = \mathit{Fix}_G(K^H)$ are equalities, for any finite extension E/k of k in K and any finite subgroup $H \subset G$.

(ii) [This is just meant to be a copy of Theorem 25.25(ii)].

The following special case of Theorem 25.29 will constitute the non-formal input into its proof, just as Lemma 25.27 did for Theorem 25.25.

Lemma 25.30. *Let G be a finite group of automorphisms of a field K, and let $k = K^G \subset K$ be the subfield of K fixed by G. Then K/k is Galois, and the obvious inclusion $G \to \mathrm{Aut}_{k\text{-}Alg}(K) = \mathrm{Gal}(K/k)$ is an isomorphism.*

Proof. We first consider the case where K/k is finite. Then by the primitive element theorem, there exists $\alpha \in K$ such that $K = k[\alpha]$. If $H \subset G$ is the stabilizer of α, set

$$f(x) = \prod_{\sigma \in [G/H]} (x - \sigma(\alpha)) \in K[x],$$

where $[G/H] \subset G$ is a set of representatives for G/H. It is easy to see that $\{\sigma(\alpha) \mid \sigma \in [G/H]\}$ is stable under G, so we have $f(x) \in (K[x])^G = k[x]$. K is clearly the splitting field of f over k, and is hence normal over k. Since f is clearly separable (the $\sigma(\alpha)$ being pairwise distinct), K/k is separable as well, and hence Galois.

Since $\deg f = \#(G/H)$, we have

$$\#G \leqslant \mathrm{Aut}_{k\text{-}Alg}(K) = \#\,\mathrm{Gal}(K/k) \leqslant [K : k]_s \leqslant [K : k] \leqslant \#(G/H),$$

forcing $\#G = \#\,\mathrm{Gal}(K/k)$, so that the obviously injective map $G \to \mathrm{Aut}_{k\text{-}Alg}(K) = \mathrm{Gal}(K/k)$ is surjective as well.

This proves the theorem when K/k is finite. To finish the proof, it suffices to show that this is the only case that arises, which will follow if we show that any finite subextension $k \subset E \subset K$ satisfies $[E : k] \leqslant \#G$. Replacing E by the compositum in K of the $\sigma(E)$ as σ varies over G (these are finitely many finite extensions, and hence their compositum in K is finite over k), we may and do assume that the action of G on K restricts to one on E. We then have $k = E^{\bar{G}}$, where $\bar{G}$ is the image of G in $\mathrm{Aut}_{k\text{-}Alg}(E)$. Therefore, by the case already handled, we have $[E : k] = \#\bar{G} \leqslant \#G$, as desired. □

Proof of Theorem 25.29. (ii) is just a repeat of Theorem 25.25(ii), so it is enough to prove (i).

Given Theorem 25.25, this follows if we show that (114) is a left-inverse to (115) as well, i.e., that given any subgroup $H \subset G$, the inclusion $H \hookrightarrow \mathrm{Fix}_G(K^H)$ is an equality. However, this is immediate from Lemma 25.30. □

25.8. **Some examples.** Our examples will mostly be taken from Serge Lang's book. For more examples, look for Keith Conrad's notes. Before we start, some generalities:

Lemma 25.31. *Let k be a field. Let $f \in k[x]$ be a separable polynomial of degree n, with a (necessarily Galois) splitting field k_f, say with (distinct) roots $\alpha_1, \ldots, \alpha_n \in k_f$.*

(i) If f is irreducible, then $[k_f : k] = \#\,\mathrm{Gal}(k_f/k)$ is a multiple of n.
(ii) For each $\sigma \in \mathrm{Gal}(k_f/k)$, there is a unique permutation $a_\sigma \in S_n$ such that $\sigma(\alpha_i) = \alpha_{a_\sigma(i)}$ for $1 \leqslant i \leqslant n$. Moreover, sending σ to a_σ defines an injective homomorphism $\mathrm{Gal}(k_f/k) \hookrightarrow S_n$, so that $[k_f : k]$ divides $n!$.

Proof. Since $[k_f : k] = [k_f : k[\alpha_1]][k[\alpha_1] : k] = n[k_f : k[\alpha_1]]$, $\#\,\mathrm{Gal}(k_f/k) = [k_f : k]$ is a multiple of n, giving the first assertion. The second assertion is immediate. □

25.8.1. *A trivial situation.* What are the field extensions between $k = \mathbb{Q}$ and $K = \mathbb{Q}[\sqrt{3}, \sqrt{7}]$? To compute this directly with elements of K may not be very pleasant, but Galois theory gives us an easy way out: it is easy to see that K/k is Galois with $\mathrm{Gal}(K/k) \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$, say generated by σ and τ with $\sigma(\sqrt{3}) = \sqrt{3}, \sigma(\sqrt{7}) = -\sqrt{7}, \tau(\sqrt{3}) = -\sqrt{3}$ and $\tau(\sqrt{7}) = \sqrt{7}$.

The main theorem of Galois theory (Theorem 25.29) tells us that the intermediate fields are in bijection with subgroups of $\mathrm{Gal}(K/k) \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. There are three nontrivial proper subgroups among these, namely those generated by σ, τ and $\sigma\tau$. It is easy to see that $K^\sigma = \mathbb{Q}[\sqrt{3}], K^\tau = \mathbb{Q}[\sqrt{7}]$ and $K^{\sigma\tau} = \mathbb{Q}[\sqrt{21}]$, so it follows that the only intermediate fields between $\mathbb{Q}$ and $\mathbb{Q}[\sqrt{3}, \sqrt{7}]$ that are neither $\mathbb{Q}$ nor $\mathbb{Q}[\sqrt{3}, \sqrt{7}]$ are $\mathbb{Q}[\sqrt{3}], \mathbb{Q}[\sqrt{7}]$ and $\mathbb{Q}[\sqrt{21}]$.

In fact, the 'easy' half, Theorem 25.25, sufficed for the above computation.

25.8.2. *The case of finite fields.* Recall that any finite field has characteristic $p > 0$ for some prime p, and then, being a vector space over $\mathbb{F}_p$, has cardinality p^n for some n. Such a field is then a splitting field of $x^{p^n-1} - 1 = 0$ over $\mathbb{F}_p$, and hence its isomorphism class depends only on its cardinality. Therefore, we may and do denote it by $\mathbb{F}_{p^n}$.

Lemma 25.32. *There exists a homomorphism $\mathbb{F}_{p^m} \to \mathbb{F}_{p^n}$ (tautologically a homomorphism of $\mathbb{F}_p$-algebras) if and only if $m|n$.*

Proof. For "$\Rightarrow$", use that such a homomorphism would make $\mathbb{F}_{p^n}$ an $\mathbb{F}_{p^m}$-vector space. "$\Leftarrow$" follows from the fact that if m divides n, then $p^m - 1$ divides $p^n - 1$, and hence $x^{p^m-1} - 1$ divides $x^{p^n-1} - 1$. □

Lemma 25.33. *Each extension of finite fields is Galois.*

Proof. Easy exercise using that $\mathbb{F}_{p^n}$ is a splitting field for $x^{p^n-1} - 1$ over any subfield. Alternatively, this is a special case of Lemma 25.34 below. □

Lemma 25.34. *Any extension K/k of finite fields is cyclic, with Galois group cyclic of order $[K : k]$, generated by* $\mathrm{Frob}_k : x \mapsto x^{\#k}$.

Proof. It is easy to see that Frob_k belongs to $\mathrm{Aut}_{k\text{-}Alg}(K/k)$ and has finite order. Further, the subfield of K fixed by $\langle \mathrm{Frob}_k \rangle$ consists of solutions to $x^{\#k} - x = 0$, and can hence be no larger than k. Therefore, it follows from Lemma 25.30 that K/k is Galois, with $\mathrm{Gal}(K/k) = \langle \mathrm{Frob}_k \rangle$ (we leave a more 'elementary' proof avoiding the use of Lemma 25.30 as an easy exercise). □

Exercise 25.35. If k is a finite field and $k \hookrightarrow \bar{k}$ is an algebraic closure, show that there exists an isomorphism $\mathrm{Gal}(\bar{k}/k) \to \hat{\mathbb{Z}}$, whose inverse sends $1 \in \mathbb{Z} \subset \hat{\mathbb{Z}}$ to $\mathrm{Frob}_k : x \mapsto x^{\#k}$.

25.8.3. *The Galois group of cubic polynomials (over characteristic $\neq 2, 3$).* Let us compute $\mathrm{Gal}(k_f/k)$, where k is a field of characteristic different from 2 and 3, and k_f is the splitting field of an irreducible cubic polynomial $f \in k[x]$. Since $\mathrm{char}\, k \neq 3$, it is easy to see that f is automatically separable.

Since $\mathrm{char}\, k \neq 3$, we can, up to a linear change of coordinates (which does not change the Galois group), write any cubic polybomial $f \in k[x]$ as $f(x) = x^3 + ax + b$. By Lemma 25.31, $[k_f : k]$ equals 3 or 6; more precisely, the same lemma tells us that $\mathrm{Gal}(k_f/k)$ equals either the alternating group A_3 or the symmetric group S_3 on 3 letters.

Lemma 25.36. *Let $\Delta = -4a^3 - 27b^2$. Then*

$$\mathrm{Gal}(k_f/k) \cong \begin{cases} A_3, & \text{if } \Delta \text{ is a square in } k, \text{ and} \\ S_3, & \text{if } \Delta \text{ is not a square in } k \end{cases}.$$

Example 25.37. Let $k = \mathbb{Q}$ and $f(x) = x^3 - x - 1$. It is easy to check that f is irreducible over $\mathbb{Q}$ (else by Gauss's lemma, it would have a root over $\mathbb{Z}$, which by prime factor considerations would be ± 1), and that $\Delta = -23$, so by the above lemma, we have $\mathrm{Gal}(k_f/k) \cong S_3$.

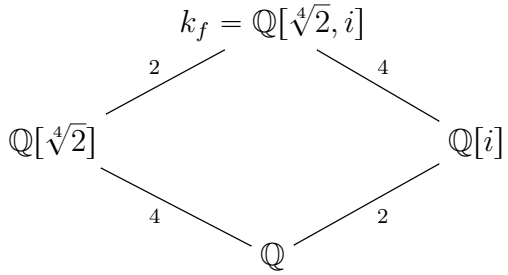
On the other hand, suppose $k = \mathbb{Q}$ and $f(x) = x^3 - 3x + 1$. Again using Gauss's lemma, one sees that f is irreducible. We have $\Delta = 81$, so by the above lemma, so $\mathrm{Gal}(k_f/k) \cong A_3$.

Proof of Lemma 25.36. We will use (exercise!) the computation that, if $\alpha_1, \alpha_2, \alpha_3 \in k_f$ are the roots of f, then setting $\delta = (\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_3 - \alpha_1)$, we have $\delta^2 = \Delta$.

Use the injection $G := \mathrm{Gal}(k_f/k) \hookrightarrow S_3$ from the proof of Lemma 25.31, to view $\mathrm{Gal}(k_f/k)$ as a subgroup of S_3. Since 3 is prime, the lemma shows that the image of $\mathrm{Gal}(k_f/k)$ contains A_3. The description of the injection $\mathrm{Gal}(k_f/k) \hookrightarrow S_3$ makes it clear that $\sigma \in G$ maps to an element of $S_3 \backslash A_3$ if and only if $\sigma(\delta) = -\delta$. Since $k_f^{\mathrm{Gal}(k_f/k)} = k$, it follows that $G \cong S_3$ if and only if $\delta \notin k$, i.e., if and only if Δ is not a square in k. □

25.8.4. *The Galois group of the polynomial $f(x) = x^4 - 2$ over $\mathbb{Q}$.* We let $k = \mathbb{Q}$ and $f(x) = x^4 - 2$. By Eisenstein's criterion for irreducibility, f is irreducible. Since $\mathrm{char}\,\mathbb{Q} = 0$, f is automatically separable, and hence k_f/k is Galois. We can write $k_f = \mathbb{Q}[\pm\sqrt[4]{2}, i]$, where $\sqrt[4]{2}$ is some choice of 4-th root of 2, [73] and i is some choice of square-root of -1.

Now we have the following diagram of field extensions, where each line between fields is marked with the degree of that extension:



To compute $G := \mathrm{Gal}(k_f/\mathbb{Q})$, one studies the two obvious subgroups that the above diagram shows us, namely, $\mathrm{Gal}(k_f/\mathbb{Q}[i])$ and $\mathrm{Gal}(k_f/\mathbb{Q}[\sqrt[4]{2}])$.

It is easy to compute that $\mathrm{Gal}(k_f/\mathbb{Q}[i]) \subset \mathrm{Gal}(k_f/\mathbb{Q})$ is cyclic of order 4, generated by a unique element σ with $\sigma(\sqrt[4]{2}) = \sqrt[4]{2}i$. On the other hand, it is easy to see that $\mathrm{Gal}(k_f/\mathbb{Q}[\sqrt[4]{2}])$ is cyclic of order 2, generated by a unique element τ with $\tau(i) = -i$. Note that $\tau\sigma\tau^{-1}$ maps $\sqrt[4]{2}$ to $-\sqrt[4]{2}i$, and i to i. Therefore, $\tau\sigma\tau^{-1} = \sigma^3 = \sigma^{-1}$.

Since it is easy to see that $\langle\sigma\rangle \cap \langle\tau\rangle = \{1\}$ inside G, we have $\#(\langle\sigma, \tau\rangle) = 8 = [k_f : \mathbb{Q}]$, forcing $\langle\sigma, \tau\rangle = G$ as well as that $G = \langle\sigma, \tau\rangle$ is the dihedral group with eight elements (since $\sigma^4 = \tau^2 = 1$ and $\tau\sigma\tau^{-1} = \sigma^{-1}$). (We will typically denote this group by D_8, though some sources also write D_4 for it).

Exercise 25.38. Write down all the subgroups $H \subset D_8 = \mathrm{Gal}(\mathbb{Q}[\sqrt[4]{2}, i]/\mathbb{Q})$, and compute the corresponding intermediate field extensions.

[73] It is not natural to consider the 'positive' fourth root of 2 here; this computation has nothing to do with $\mathbb{R}$.

25.8.5. *'Elementary symmetric rational functions'.* Let F be a field, and let $K = F(t_1, \ldots, t_n)$, the field of rational functions in n variables $t_1, \ldots, t_n$ over F (which is the quotient field of the polynomial ring $F[t_1, \ldots, t_n]$). Let $s_1, \ldots, s_n \in F[t_1, \ldots, t_n] \subset K$ be the elementary symmetric polynomials, so that

$$s_i = \sum_{1 \leqslant j_1 < j_2 < \cdots < j_i \leqslant n} t_{j_1} t_{j_2} \ldots t_{j_i}.$$

Set $k = F(s_1, \ldots, s_n) \subset K$.

Note that $G = S_n$ acts on K, fixing F pointwise and with σ sending t_i to $t_{\sigma(i)}$ for each i.

Lemma 25.39. *$k = K^G$, so that K/k is Galois with Galois group isomorphic to S_n.*

Proof. Clearly $k \subset K^G$. By Lemma 25.30, we have

$$n! = \#G = [K : K^G] \leqslant [K : k].$$

Therefore, it suffices to show that $[K : k] \leqslant n!$. Note that $t_1, \ldots, t_n$ are roots of the polynomial $f \in k[x]$, where $f(x) = x^n + \sum_{i=1}^{n} (-1)^i s_i x^{n-i}$. Since f is separable, K/k is Galois. By Lemma 25.31, $[K : k] = \#\operatorname{Gal}(K/k)$ divides $n!$, as desired. □

From this, it follows that any $f \in k[t_1, \ldots, t_n]$ that is invariant under the action of G belongs to $k(s_1, \ldots, s_n) \cap k[t_1, \ldots, t_n]$. In fact, it follows from some standard facts that $k(s_1, \ldots, s_n) \cap k[t_1, \ldots, t_n] = k[s_1, \ldots, s_n]$. Thus, one can show that any polynomial in $k[t_1, \ldots, t_n]$ invariant under the action of S_n is a polynomial in the elementary symmetric polynomials, $s_1, \ldots, s_n$.

Exercise 25.40. Use the above lemma to show that given any finite group G, there is a Galois extension K/k of fields such that $\operatorname{Gal}(K/k) \cong G$. However, it is an open problem as to whether there exists such an extension with $k = \mathbb{Q}$.

25.8.6. *A Galois-theoretic proof of the fundamental theorem of algebra.*

Theorem 25.41. *$\mathbb{C}$ is algebraically closed.*

Proof. Suppose $\mathbb{C}$ is not algebraically closed. Then $\mathbb{R}$ has a Galois extension K properly containing $\mathbb{C}$. Let $G = \operatorname{Gal}(K/\mathbb{R})$. Let $H \subset G$ be a 2-Sylow subgroup. By the main theorem of Galois theory (Theorem 25.29), there exists an intermediate field extension $\mathbb{R} \subset F \subset K$ such that $\operatorname{Gal}(K/F) \cong H$. But this implies that $[F : \mathbb{R}] = \#(G/H)$ is odd.

We claim that $F = \mathbb{R}$. Suppose not. Then there exists $\alpha \in F \backslash \mathbb{R}$ such that the minimal polynomial of α over $\mathbb{R}$ has odd degree. Since such a minimal polynomial being irreducible over $\mathbb{R}$ has no root in $\mathbb{R}$, this contradicts the fact that any odd degree polynomial $f \in \mathbb{R}[x]$ has a root in $\mathbb{R}$ (proof without using fundamental theorem of algebra: w.l.o.g. f is monic, so $f(x) \to -\infty$ as $x \to -\infty$ and $f(x) \to \infty$ as $x \to \infty$, so f has a real root by the intermediate value theorem). This contradiction proves that $F = \mathbb{R}$.

Therefore, $[K : \mathbb{R}]$ is a power of two, and hence so is $[K : \mathbb{C}]$. Since $K/\mathbb{R}$ is Galois, so is $K/\mathbb{C}$. It is enough to show that $K = \mathbb{C}$. Suppose not.

Since any 2-group is nilpotent, it is easy to see that there is a surjection $\mathrm{Gal}(K/\mathbb{C}) \to \mathbb{Z}/2\mathbb{Z}$, which gives a nontrivial quadratic extension of $\mathbb{C}$ (again, use the main theorem of Galois theory, Theorem 25.29). However, any quadratic extension of $\mathbb{C}$ is obtained by adjoining a root of an irreducible quadratic polynomial to $\mathbb{C}$. This is a contradiction, since by the quadratic formula, any quadratic polynomial over $\mathbb{C}$ is reducible. □

Remark 25.42. According to Serge Lang's book, the ideas in the proof above were already there in the work of Gauss, but Artin made it efficient using Sylow subgroups etc.

25.8.7. *Finding a polynomial in $\mathbb{Q}[x]$ with Galois group S_5.*

Lemma 25.43. *Let p be a prime number, and let $f \in \mathbb{Q}[x]$ be irreducible of degree p. Assume that f has precisely two non-real roots in $\mathbb{C}$. Then $\mathrm{Gal}(\mathbb{Q}_f/\mathbb{Q}) \cong S_p$, where $\mathbb{Q}_f$ is a splitting field of $\mathbb{Q}$.*

Proof. Fix a $\mathbb{Q}$-algebra embedding $\mathbb{Q}_f \hookrightarrow \mathbb{C}$, and think of it as an inclusion. Lemma 25.31 gives an embedding $G \hookrightarrow S_p$, with G acting by permuting the roots of f in $\mathbb{Q}_f \subset \mathbb{C}$. Complex conjugation preserves $\mathbb{Q}_f \subset \mathbb{C}$ (since $\mathbb{Q}_f$ is a splitting field of f in $\mathbb{C}$), and gives an element of G whose image in S_p permutes the two non-real roots of f in $\mathbb{C}$ and fixes the rest. Thus, the image of G in S_p contains a p-cycle and a transposition, and is hence the whole of S_p (since any p-cycle and any transposition together generate S_p). □

Example 25.44. Consider $f \in \mathbb{Q}[x]$ given by $f(x) = x^5 - 4x + 2$. By Eisenstein's criterion for irreducibility, f is irreducible. We claim that the Galois group of f is isomorphic to S_5. By Lemma 25.43 above, this follows if we show that f has exactly two non-real roots, which we claim to be the case.

Since $f'(x) = 4(x^4 - 1)$, f is increasing in $(-\infty, -1] \cup [1, \infty)$ and decreasing in $[-1, 1]$. Since $f(-1) > 0$ and $f(1) < 0$, it follows that f has exactly one root each in each of $(-\infty, -1), (-1, 1)$ and $(1, \infty)$. Thus, f has exactly two non-real roots, as desired.

26. Lecture 26 – Galois theory

Note: This lecture was the most experimental of the lecture series, and hence the notes for this lecture are much more likely to have serious errors than the notes for other lecture.

In Lecture 25 we saw a classical proof of the main theorem of Galois theory for finite Galois extensions – in its the classical 'bijection' version rather than the 'equivalence of categories' version. Today, we will, among other things, discuss what I believe Artin was getting at in the proof he gave in his Notredame lectures on Galois theory, namely, Galois theory via Galois descent. Thus, rather than proving Galois theory and using it to prove Galois descent, we will prove Galois descent and use it to prove an 'equivalence of categories' version of the Galois correspondence. For this, though, will assume the 'easy half' of Galois theory, which says that if K/k is a (possibly infinite) Galois extension, then $K^{\mathrm{Gal}(K/k)} = k$ – see Lemma 25.27 from Lecture 25. This half only used basic facts about separability and about extending homomorphisms into an algebraically closed field, and did not use the primitive element theorem. We will also discuss another take on the main theorem of Galois theory, this time for finite extensions, which again avoids the use of primitive element theorem, and then the infinite Galois correspondence.

26.1. **Galois descent.** First, some very informal motivation. Given a vector space $V_{\mathbb{R}}$ over $\mathbb{R}$, we have a vector space $V_{\mathbb{C}} := \mathbb{C} \otimes_{\mathbb{R}} V_{\mathbb{R}}$ over $\mathbb{C}$: this is base-change/extension of scalars. We have an inclusion $V_{\mathbb{R}} \hookrightarrow V_{\mathbb{C}}$ that sends each w to $1 \otimes w$, letting us view $V_{\mathbb{R}}$ as an $\mathbb{R}$-subspace of $V_{\mathbb{C}}$.

How does one go in the opposite direction, to recover $V_{\mathbb{R}}$ from$V_{\mathbb{C}}$ (plus extra data)? Note that there is an $\mathbb{R}$-linear operator $v \mapsto \bar{v}$ on $V_{\mathbb{C}}$, sending $a \otimes w$ to $\bar{a} \otimes w$ for each $w \in V_{\mathbb{R}}$ and $a \in \mathbb{C}$, where $\bar{a}$ is the complex conjugate of a. It is clear that $V_{\mathbb{R}} \subset V_{\mathbb{C}}$ is simply the subspace fixed by $v \mapsto \bar{v}$. In other words, $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ acts on $\mathbb{C} \otimes_{\mathbb{R}} V_{\mathbb{R}} = V_{\mathbb{C}}$ through its action on $\mathbb{C}$, and $V_{\mathbb{R}} = V_{\mathbb{C}}^{\mathrm{Gal}(\mathbb{C}/\mathbb{R})}$ is the fixed subspace. Note that this action is not $\mathbb{C}$-linear, but is $\mathbb{R}$-linear, and is in fact what one calls $\mathbb{C}$-semilinear: $\overline{a \cdot v} = \bar{a} \cdot v$ for all $a \in \mathbb{C}$ and $v \in V_{\mathbb{C}}$.

To summarize, $V_{\mathbb{R}}$ determines both $V_{\mathbb{C}}$ and a semilinear $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$-action $v \mapsto \bar{v}$ on $V_{\mathbb{C}}$, and these two data together get us $V_{\mathbb{R}}$ back. Something similar can be done with $\mathbb{C}/\mathbb{R}$ replaced by any Galois extension (with extra care in the case of infinite Galois extensions), and this is what Galois descent is.

Definition 26.1. Let K/k be an arbitrary Galois extension, and let $G = \mathrm{Gal}(K/k)$. A k-linear action of G on a K-vector space V is said to be:

(i) (K-)semilinear, if $\sigma(a \cdot v) = \sigma(a) \cdot \sigma(v)$ for all $a \in K$ and $v \in V$ (thus, such an action is not K-linear); and
(ii) continuous, if for all $v \in V$, there exists a finite subextension F/k of K/k such that $\mathrm{Gal}(K/F) \subset \mathrm{Gal}(K/k)$ fixes v; in other words,

$$V = \bigcup_H V^H, \tag{116}$$

where H runs over the $\mathrm{Gal}(K/F)$ as F runs over the finite extensions of k contained in K.

Remark 26.2. In Exercise 26.27 below, we will see that this notion of continuity is equivalent to the continuity of the action map $\mathrm{Gal}(K/k) \times V \to V$, where V is given the discrete topology, and $\mathrm{Gal}(K/k)$ is given the Krull topology defined in Definition 26.25 below.

Example 26.3. If V_k is a vector space over K, then $G = \mathrm{Gal}(K/k)$ acts on $V := K \otimes_k V_k$ such that $\sigma(a \otimes v) = \sigma(a) \otimes v$ for all $\sigma \in G, a \in K$ and $v \in V_k$. This action is easily checked to be both semilinear and continuous – continuous because each $a \in K \backslash k$, being algebraic over k, is contained in a finite subextension F/k of K/k, and is hence fixed by $\mathrm{Gal}(K/F)$.

Theorem 26.4. *Let K/k be Galois, and let $G := \mathrm{Gal}(K/k)$. Sending $V_k \in \mathrm{Ob}\, Vec_k$ to $V := K \otimes_k V_k$, viewed with the obvious semilinear G-action on it (Example 26.3), induces an equivalence of categories:* [74]

$$Vec_k \rightsquigarrow \{\textit{Vector spaces over } K \textit{ + continuous semilinear } G\textit{-action}\}. \tag{117}$$

It has a quasi-inverse that sends V to the k-vector space V^G (and restriction to V^G at the level of morphisms).

Exercise 26.5. Check that the functor described in Theorem 26.4 is fully faithful. (We will not need this exercise, but it is still good to see this directly).
Hint: This is very easy: it just says (the infinite dimensional version of): if a matrix with entries in K is fixed by G, then it has entries in k.

Proof of Theorem 26.4. Note that the G-action on $K \otimes_k V_k$ is indeed continuous and semilinear as discussed in Example 26.3.

Using a k-basis for V_k, and the fact that the inclusion $k \hookrightarrow K^{\mathrm{Gal}(K/k)}$ is an equality (Lemma 25.27 from Lecture 25), it is easy to see that $(K \otimes_k V_k)^{\mathrm{Gal}(K/k)}$ is simply the k-subspace $V_k \cong \{1\} \otimes_k V_k \subset K \otimes_k V_k$. [75] From this, it is immediate that the composite $(V \rightsquigarrow V^G) \circ (V_k \rightsquigarrow K \otimes_k V_k)$ is naturally isomorphic to the identity functor on Vec_k, via the natural isomorphism $((V_k \to \{1\} \otimes V_k)_{V_k}$ given on each V_k by $w \mapsto 1 \otimes w$.

In the other direction, there is a natural transformation from $(V_k \rightsquigarrow K \otimes_k V_k) \circ (V \rightsquigarrow V^G)$ to the identity functor, namely the obvious map

$$K \otimes_k V^G \to V \tag{118}$$

(i.e., with the property that $a \otimes w \mapsto aw$, namely, induced from the bilinear map $(a, w) \mapsto aw$; also, please make sure to check that these maps constitute a natural transformation). Thus, it remains to show that (118) is an isomorphism for every vector space V over K with a continuous semi-linear action of G. This is outsourced to Proposition 26.6 below. □

[74] As usual, fill it in at the level of morphisms.

[75] In more detail: if $w_1, \ldots, w_n \in V_k$ are linearly independent, $a_1, \ldots, a_n \in K$ and $(\sum a_i \otimes w_i) \in (K \otimes_k V_k)^{\mathrm{Gal}(K/k)}$, then for each $\sigma \in \mathrm{Gal}(K/k)$, we have $\sum \sigma(a_i) \otimes w_i = \sum a_i \otimes w_i$, so $\sigma(a_i) = a_i$ for each i, using the fact that $1 \otimes w_1, \ldots, 1 \otimes w_n$ are linearly independent in $K \otimes_k V_k$.

Proposition 26.6. *Let K/k be Galois, and let $G := \mathrm{Gal}(K/k)$. If V/K is a vector space equipped with a continuous semilinear action of $\mathrm{Gal}(K/k)$, then the map*

$$K \otimes_k V^G \to V$$

is an isomorphism of vector spaces.

The injectivity assertion in the proposition will be an easy consequence of the following lemma; it follows Theorem 14 of Artin's Notredame lecture notes on Galois theory, and is used in standard references on Galois descent. [76]

Lemma 26.7. *Let G be a (not necessarily finite) group acting by automorphisms on a field k, and let $k = K^G$ be the fixed field. Let V_k be a vector space over k. Consider the K-vector space $V = K \otimes_k V_k$, on which G acts via its action on K. Suppose $W \subset V$ is a nonzero G-invariant subspace. Then the k-vector space $W^G \subset W$ is nonzero.*

Proof. Suppose $W^G \neq 0$. Write a nonzero element of W^G as $w = \sum_{i=1}^r a_i v_i$, with $v_1, \ldots, v_r \in V_k$ linearly independent and $a_1, \ldots, a_r \in K$, where we may and do assume that r is minimal possible as w and its expansions as above are allowed to vary.

Without loss of generality, we may assume that $a_1 = 1$. We claim that $r = 1$ and that hence $w = v_1$ is the required vector. Suppose not. If all the a_i belong to k, then $w \in V_k$ and hence $r = 1$. So assume without loss of generality that $a_2 \notin k$. There exists $\sigma \in G$ such that $\sigma(a_2) \neq a_2$. Since W is G-invariant, it also contains $\sigma(w) = \sum_{i=1}^r \sigma(a_i) v_i$. Then $0 \neq \sigma(w) - w \in W$ is a K-linear combination of $v_2, \ldots, v_r$, contradicting the minimality of r. □

Proof of Proposition 26.6. Injectivity. Let $W = \ker(K \otimes_k V^G \to V)$. If $W \neq 0$, Lemma 26.7 shows that $W^G \neq 0$. However, $W^G = W \cap (K \otimes_k V^G)^G$ is contained in V^G, on which $K \otimes_k V^G \to V$ simply restricts to the inclusion $V^G \hookrightarrow V$, forcing $W^G \subset \ker(V^G \hookrightarrow V) = 0$, a contradiction.

Surjectivity, when K/k is finite. This argument seems similar to that in Theorem 13 of Artin's Notredame notes. Suppose $K \otimes_k V^G \to V$ is not surjective, i.e., the inclusion $\mathrm{Span}_K(V^G) \subset V$ is proper. Then there exists a K-linear functional $\lambda : V \to K$ such that $\lambda(V^G) = 0$. Fix $v \in V$ such that $\lambda(v) \neq 0$.

Since K/k is finite, $\#G \leqslant [K : k]_s < \infty$. Consider the map $K \to K$ given by

$$K \ni a \mapsto \lambda\Big(\sum_{\sigma \in G} \sigma(av)\Big) = \sum_{\sigma \in G} \lambda(\sigma(av)) = \sum_{\sigma \in G} \lambda(\sigma(a)\sigma(v)) = \sum_{\sigma \in G} \lambda(\sigma(v)) \cdot \sigma(a) \in K.$$

[76] Two key theorems in Artin's notes are Theorems 13 and 14 – Theorem 13 is where he uses the linear independence of characters, while in Theorem 14 he crucially uses that the collection of automorphisms being considered in that theorem form a group, something that is therefore nontrivially used in the proof of Lemma 26.7 as well. Theorem 14 looks like an injectivity assertion and Theorem 13 looks like a surjectivity assertion.

This map is a K-linear combination of the various $\sigma \in G$. This linear combination is nontrivial, because $\lambda(v) \neq 0$. Therefore, by the linear independence of characters, this map is nonzero. Hence, for some $a \in K$, λ does not vanish on $\sum_{\sigma \in G} \sigma(av) \in V^G$, a contradiction.

Surjectivity, general case. This time the continuity becomes relevant. By the definition of continuity (specifically, use (116)), it is enough to show that the image of $K \otimes_k V^G \to V$ contains V^H whenever $H = \mathrm{Gal}(K/F)$ for some finite extension F/k inside K. Without loss of generality, we may assume that F/k is normal and hence Galois (replace F with the splitting field of the collection of minimal polynomials of the elements from some set of generators of F over k). Note that V^H is a vector space over F (by the semilinearity assumption), and gets a semilinear action of $\mathrm{Gal}(K/k)/\mathrm{Gal}(K/F) \cong \mathrm{Gal}(F/k)$.

Moreover, $(V^H)^{\mathrm{Gal}(F/k)} = V^G$, and the map $K \otimes_k V^G \to V$ restricts to the analogous map $F \otimes_k (V^H)^{\mathrm{Gal}(F/k)} \to V^H$. Therefore, the image of $K \otimes_k V^G$ contains that of $F \otimes_k (V^H)^{\mathrm{Gal}(F/k)}$, which is V^H by the finite case that has already been handled. □

Corollary 26.8. *(Galois descent for algebras). Let K/k be Galois, and let $G := \mathrm{Gal}(K/k)$. Sending $A_k \in \mathrm{Ob}\, k\text{-}Alg$ to $A := K \otimes_k A_k$, viewed with the obvious semilinear G-action on it, induces an equivalence of categories:*

$$k\text{-}Alg \rightsquigarrow \{K\text{-algebras + continuous semilinear } G\text{-action}\}. \tag{119}$$

It has a quasi-inverse that sends A to the k-algebra space A^G (and restriction to A^G at the level of morphisms). All this holds upon adding the adjectives 'commutative', 'finite' and/or 'finite commutative'.

Proof. Just as in the proof of Theorem 26.4, we have the obvious map $A_k \to (K \otimes_k A_k)^G$ of k-algebras and the obvious map $K \otimes_k A^G \to A$ of K-algebras. It is enough to show that these maps are isomorphisms, which follows if we show that they are isomorphisms of vector spaces over k or K. This has been done in Proposition 26.6. □

Remark 26.9. Perhaps some might find it more natural, and others more pretentious, to deduce the descent for algebras from that for vector spaces by observing that the prescription for vector spaces respects the additional structure involved in the definition of an algebra A, namely, the multiplication map $A \otimes A \to A$, together with the associativity and the existence of an identity element. Of course, one needs to carefully make sense of such structures in the presence of semilinear actions, but all that is ultimately easy.

26.2. **An equivalence of categories version of the Galois correspondence.**

Definition 26.10. (i) Let K/k be a field extension. A finite commutative k-algebra A/k is said to split over K, or K-split, if $A \otimes_k K$ is a product of copies of K. Let $spl_K(k)_f$ denote the category of finite commutative k-algebras that split over K, and k-algebra homomorphisms between them. [77] If we say 'finite split', we will mean 'finite split commutative'.

[77] I understand that the notation $spl_K(k)_f$ sits in less than desirable level of harmony with the notation $spl(k)_f$ from Lecture 25, and regret that.

(ii) Let K/k be a Galois extension. The action of $G = \mathrm{Gal}(K/k)$ on a finite set (or more generally a discrete topological space) X is said to be continuous if the stabilizer of each $x \in X$ contains $\mathrm{Gal}(K/F) \subset \mathrm{Gal}(K/k)$ for a finite subextension F/k of K/k. In this case, we will write $(G\text{-}FinSet)_{cts}$ for the category of finite sets with a continuous G-action.

(iii) For each $A \in \mathrm{Ob}\, spl_K(k)_f$, $X_A := \mathrm{Hom}_{k\text{-}Alg}(A, K)$ will be viewed as a G-set, where G acts (on the left) via its action on K. It is immediate that this action is continuous, and we get a functor $spl_K(k)_f \rightsquigarrow (G\text{-}FinSet)_{cts}$, which we will denote by $A \rightsquigarrow X_A$.

(iv) For a finite set X with a continuous G-action, we will, like in Lecture 25, view the K-algebra $\mathrm{Maps}(X, K)$ as having the action of G given by $\sigma \cdot \varphi = \sigma \circ \varphi \circ \sigma^{-1}$. We get a functor $(G\text{-}FinSet)_{cts} \rightsquigarrow k\text{-}Alg$, given by $X \rightsquigarrow A_X := \mathrm{Maps}_G(X, K) := \mathrm{Maps}(X, K)^G$. It is an easy exercise to see that each A_X is finite commutative. It will be a consequence of Theorem 26.12 below that this functor is actually valued in $spl_K(k)_f$.

Lemma 26.11. *If $A \in \mathrm{Ob}(k\text{-}Alg)^{fc}$ splits over some field extension K/k, then A is separable over k, i.e., $A \in \mathrm{Ob}\, fét_k$. It follows that $spl_K(k)_f$ is a (full by definition) subcategory of $fét_k$.*

Proof. We saw in Lecture 24 that A is separable over k if and only if $K \otimes_k A$ is separable over K. Now use that finite split K-algebras are separable over K. □

The 'equivalence of categories' version of the main theorem of Galois theory for finite extensions is:

Theorem 26.12. *Let K/k be a Galois extension, and set $G = \mathrm{Gal}(K/k)$. There exists an equivalence of categories*

$$spl_K(k)_f \rightsquigarrow (G\text{-}FinSet)_{cts}, \qquad A \mapsto X_A = \mathrm{Hom}_{k\text{-}Alg}(A, K),$$ [78]

with a quasi-inverse given by

$$X \rightsquigarrow A_X = \mathrm{Maps}(X, K)^G = \mathrm{Maps}_G(X, K).$$

More precisely, the composites in either direction have natural isomorphisms with the identity functors $\mathrm{id}_{spl_K(k)_f}$ and $\mathrm{id}_{(G\text{-}FinSet)_{cts}}$, given by

$$Gelf : A \to \mathrm{Maps}(X_A, K)^G, \qquad a \mapsto (\varphi \mapsto \varphi(a)), \tag{120}$$

and

$$X \mapsto \mathrm{Hom}_{k\text{-}Alg}(A_X, K), \qquad \varphi \mapsto (a \mapsto \varphi(a)). \tag{121}$$

Remark 26.13. (i) Thus, again the functors in either direction are obtained from the G-equivariant evaluation pairing

$$ev : A \times \mathrm{Hom}_{k\text{-}Alg}(A, K) \to K,$$

[78] We are changing notation here: now A denotes a k-algebra, and $K \otimes_k A$ will denote the associated K-algebra.

where A has the trivial action of G.

(ii) Taking $k \hookrightarrow K$ to be a separable closure $k \hookrightarrow k^s$, this theorem specializes to (or rather, makes precise) Theorem 25.14 from Lecture 25.

Proof of Theorem 26.12. By the definition of $spl_K(k)_f$, the equivalence of categories (119) induces an equivalence of categories:

$$spl_K(k)_f \rightsquigarrow \{\text{finite split } K\text{-algebras + continuous semilinear } G\text{-action}\}. \tag{122}$$

Recall that this sends A to $K \otimes_k A$, and has a quasi-inverse that sends B to $B^{\mathrm{Gal}(K/k)}$.

On the other hand, recall from Exercise 25.9 of Lecture 25, that the category $spl(K)_f = spl_K(K)_f$ of finite split K-algebras is equivalent to the category of finite sets, by the functor $B \rightsquigarrow X_B = \mathrm{Hom}_{K\text{-}Alg}(B, K)$, which has a quasi-inverse given by $X \rightsquigarrow B_X := \mathrm{Maps}(X, K)$. More precisely, the obvious map $B \to B_{X_B}$ is a K-algebra isomorphism, and the obvious map $X \to X_{B_X}$ is a bijection.

If B additionally has a continuous semilinear G-action, then $X_B = \mathrm{Hom}_{K\text{-}Alg}(B, K)$ gets a continuous G-action, provided for $\sigma \in G$ and $\varphi \in X_B$ we define $\sigma \cdot \varphi = \sigma \circ \varphi \circ \sigma^{-1}$ (check that $\sigma \circ \varphi \circ \sigma^{-1}$ is a K-algebra homomorphism). Similarly, if X has a continuous G-action, then B_X gets a G-action, by $\sigma \cdot (b : X \to K) = (\sigma \circ b \circ \sigma^{-1} : X \to K)$; this action is readily verified to be continuous and semilinear.

It is immediately verified that the maps $B \to B_{X_B}$ and $X \to X_{B_X}$ respect G-actions, and are hence isomorphisms respectively in the category of finite split K-algebras with a continuous semilinear G-action, and the category $(G\text{-}FinSet)_{cts}$ (since they respect G-actions, we can ignore the G-action in checking that they are isomorphisms). Therefore, we get an equivalence of categories:

$$\{\text{Finite split } K\text{-algebras + continuous semilinear } G\text{-action}\} \rightsquigarrow (G\text{-}FinSet)_{cts}. \tag{123}$$

Combining (122) and (123), we get an equivalence of categories

$$spl_K(k)_f \rightsquigarrow (G\text{-}FinSet)_{cts},$$

given by

$$A \rightsquigarrow \mathrm{Hom}_{K\text{-}Alg}(K \otimes_k A, K) = \mathrm{Hom}_{k\text{-}Alg}(A, K)$$

(use Hom-tensor adjointness), and with a quasi-inverse given by

$$X \mapsto \mathrm{Maps}(X, K)^G = \mathrm{Maps}_G(X, K),$$

as desired. □

To relate the above 'equivalence of categories' version of the Galois correspondence with the 'classical version', we will also need:

Proposition 26.14. *Let K/k be a Galois extension, and let $G = \mathrm{Gal}(K/k)$. Let F be a finite K-algebra admitting a k-algebra homomorphism $F \to K$, and let $X_F = \mathrm{Hom}_{k\text{-}Alg}(F, K)$. Then*

(i) We have an isomorphism of K-algebras $K \otimes_k F \to \mathrm{Maps}(X_F, K)$, which satisfies $b \otimes a \mapsto (\sigma \mapsto b\sigma(a))$.
(ii) $F \in \mathrm{Ob}\, spl_K(k)_f$.

Proof, some verifications only sketched. If we knew (ii), (i) would follow from the equivalence of categories between finite split K-algebras and finite sets (Exercise 25.9 from Lecture 25). The point is that we are instead using Galois descent to prove (i), and then deducing (ii).

For (i), we fix a k-algebra homomorphism $F \to K$ and view it as an inclusion. It is easy to see from the defining properties of the tensor product that there exists a well-defined K-algebra homomorphism $K \otimes_k F \to \mathrm{Maps}(X_F, K)$ satisfying $b \otimes a \mapsto (\sigma \mapsto \sigma(a)b)$. This homomorphism respects the obvious continuous semilinear actions of $G := \mathrm{Gal}(K/k)$ on either side (this needs a little bit of straightforward checking; for more details, follow Remark 26.16 below). Therefore, by Galois descent (Theorem 26.4), (i) follows if we show that this map becomes an isomorphism on taking G-invariants. This is easy, once one notes that, since G acts transitively on $\mathrm{Hom}_{k\text{-}Alg}(F, K)$, evaluating at the inclusion $F \hookrightarrow K$ (whose stabilizer in G is $\mathrm{Gal}(K/F)$) identifies $\mathrm{Maps}_G(X_F, K)$ with $K^{\mathrm{Gal}(K/F)} = F$.

(ii) follows from (i). □

26.3. **Another take on the main theorem.**
Note: The material of this section seems to give a proof of the main theorem of Galois theory, in the case of finite extensions, without any primitive element theorem or Galois descent, linear independence of characters etc. So it might be erroneous, be cautious.

In this subsection, we will give yet another proof of the classical form of the Galois correspondence for finite Galois extensions K/k. This will crucially use Exercise 25.9 of Lecture 25 (which was already used above), so we restate it as a proposition:

Proposition 26.15. *Let G be a (not necessarily finite) group acting by automorphisms on a field K, and let $k = K^G \subset K$ be the fixed field. Assume that K is algebraic over k. Let $A \in \mathrm{Ob}\, spl_K(k)_f$. Then the map*

$$K \otimes_k A \to \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A,K)} K = \mathrm{Maps}(X_A, K), \tag{124}$$

sending each $b \otimes a$ to $(b\sigma(a))_\sigma = (\sigma \mapsto b\sigma(a))$, [79] *is an isomorphism of k-algebras (thus, explicitly splitting the k-algebra K), respecting the obvious action of G on either side.*

Remark 26.16. For later use, since (124) is important, we study its equivariance properties: [80]

[79] Again, it implicitly has been left to you to check that this definition makes sense. e.g., the σ-th factor is obtained using the universal property of tensor products of commutative k-algebras, from $\mathrm{id}_K : K \to K$ and $\sigma : A \to K$.

[80] Apologies again for the bad notation: this involves the switching of some orders in maps, necessitating more care on the part of the reader than should have been demanded.

(i) First, consider the case where K is finite and $A = K$: taking $A = K$ is justified, since the k-algebra K is K-split, by Lemma 26.17 below.

In this case, we are studying $K \otimes_k K \to \prod_{\sigma \in G} K$, where $G = \mathrm{Gal}(K/k)$ is finite since A/k is assumed to be. Since the image of $b \otimes a$ in $\mathrm{Maps}(G, K)$ is given by $\sigma \mapsto c_\sigma := b\sigma(a)$, and since

$$\tau_1(b)\sigma(\tau_2(a)) = \tau_1(b \cdot (\tau_1^{-1}\sigma\tau_2(a))) = \tau_1(c_{\tau_1^{-1}\sigma\tau_2}),$$

it follows that the action of $\tau_2 \in G \cong \{1\} \times G$ on $K \otimes_k K$ gets transported to the right-regular action of $\tau_2 \in G$ on $\mathrm{Maps}(G, K)$, while the action of $\tau_1 \in G \cong G \times \{1\}$ gets transported to the action of τ_1 on $\mathrm{Maps}(G, K)$ that results from viewing G and K as having their usual left G-actions.

(ii) If we drop the assumption that $A = K$, then we still have an action of $G \cong G \times \{1\}$ on $K \otimes_k A$, which by the same computation is obtained by viewing G and K as having their usual left G-actions, and giving $\mathrm{Maps}(G, K)$ the resulting G-action.

Proposition 26.17. *If K/k is Galois, then finite K-split k-algebras are precisely those isomorphic to products of the form $\prod_{i=1}^n F_i$, with each F_i/k a finite subextension of K/k.*

Proof. Suppose A is a finite K-split k-algebra. Then A is automatically separable over k (Lemma 26.11), so we can write $A = \prod_{i=1}^n F_i$, with each F_i/k a finite separable extension. Since $K \otimes_k A$ is a product of copies of K, for each i, we have a nonzero homomorphism $K \otimes_k F_i \to K$, so that we have a k-algebra embedding $F_i \to K$. This shows that each finite K-split k-algebra is a product $\prod_{i=1}^n F_i$ as given.

We now need to prove the converse, i.e., that each finite subextension F/k of K/k is K-split (compare with Proposition 26.14). If we can assume the primitive element theorem, so that $F = k[\alpha]$, then the minimal monic polynomial f of α over k has a factorization of the form $(x - \alpha_1) \dots (x - \alpha_n)$, with the α_i distinct, so Sunzi's theorem gives:

$$K \otimes_k F \cong K[x]/\Big(\prod_{i=1}^n (x - \alpha_i)\Big) \cong \prod_{i=1}^n K,$$

as a K-algebra, which is split.

But we wish to prove the proposition without using the primitive element theorem, since we wish to have a proof of the main theorem of Galois theory without using the primitive element theorem. The above argument does show that for any $\alpha \in F$, $K \otimes_k k[\alpha] \subset K \otimes_k F$ is a split K-algebra. It follows that $K \otimes_k F$ is generated as a K-subalgebra by split K-subalgebras. Thus, using finiteness of F/k, it is enough to show that split subalgebras of $K \otimes_k F$ are closed under taking compositums. This is because the compositum of K-subalgebras $A_1, \dots, A_n$ of a K-algebra A is a homomorphic image of $A_1 \otimes_K \dots \otimes_K A_n$, and it is easy to see that finite split K-algebras are closed under taking tensor products as well as (ring-theoretic) quotients. □

Another take on the "$Fix(K^H) = H$" half of Galois theory, when K/k is finite. For $H \subset G = \mathrm{Gal}(K/k)$, we need to show that the inclusion $H \subset H'$ is an equality, where $H' =$

$Fix_G(K^H)$ is the fixer of K^H in G. From the 'easy half' of Galois theory, we know that $\#H' = \#\operatorname{Gal}(K/K^H) = [K : K^H]$, so it is enough to show that $\#H = [K : K^H]$.

We apply Proposition 26.15 with $A = K$ – we can indeed take $A = K$, by Proposition 26.17. Note that $\operatorname{Hom}_{k\text{-}Alg}(A, K) = \operatorname{Hom}_{k\text{-}Alg}(K, K) = G$. For $H \subset G$, taking H-fixed points and using Remark 26.16(ii), we get an isomorphism $K^H \otimes_k K \to \operatorname{Maps}_H(G, K)$ of k-algebras (even of K^H-algebras). The dimension of $K^H \otimes_k K$ over k is $[K^H : k] \cdot [K : k]$, while, since H acts on G without fixed points, the dimension of $\operatorname{Maps}_H(G, K)$ over k is $\#(G/H) \cdot [K : k]$. This gives $\#(G/H) = [K^H : k]$, and since $\#G = [K : k]$, we get $[K : K^H] = \#H$, as desired. □

Remark 26.18. Perhaps it would be good to give a proof of the 'equivalence of categories' version using these ideas, but I have not thought about how to do that.

26.4. **The relation between the classical version and the 'equivalence of categories' version.** Let us state a classical form of the main assertion of the main theorem of Galois theory, describing finite subextensions of a possibly infinite Galois extension K/k (the case where K/k is finite was covered in Lecture 25): [81]

Theorem 26.19. *Let K/k be a (not necessarily finite) Galois extension. Then we have a bijection*

$$\left\{\textit{Finite subextensions } F/k \textit{ of } K/k\right\} \to \left\{\begin{smallmatrix}\textit{Subgroups of } \operatorname{Gal}(K/k) \textit{ containing} \\ \operatorname{Gal}(K/F_0) \textit{ for some finite subextension } F_0/k\end{smallmatrix}\right\}, \quad F \mapsto H_F := Fix_G(F) = \operatorname{Gal}(K/F), \tag{125}$$

with a two-sided inverse

$$H \mapsto K^H. \tag{126}$$

Proof. As an exercise, prove this using the classical form of the main theorem of Galois theory that we saw in Lecture 25. Let us instead prove it using the 'equivalence of categories' version, thereby explicating the relation between the 'equivalence of categories' version and the classical version.

The 'equivalence of categories' version of the main theorem of Galois theory (Theorem 26.12) says that the following two obvious maps are isomorphisms in the appropriate category (see (120) and (121)):

$$A \to \operatorname{Maps}(X_A, K)^G, \tag{127}$$

$$X \mapsto \operatorname{Hom}_{k\text{-}Alg}(A_X, K). \tag{128}$$

By Proposition 26.17, the condition that the maps (127) are all isomorphisms needs to be checked only on finite subextensions F/k of K/k. For such an F, $X_A = \operatorname{Hom}_{k\text{-}Alg}(F, K)$ identifies with $\operatorname{Gal}(K/k)/\operatorname{Gal}(K/F)$, so the map $A \to \operatorname{Maps}(X_A, K)^G$ identifies with $F \to$

[81] The right-hand side will be described in terms of a topology in Theorem 26.26 below.

$K^{H_F} = K^{\mathrm{Gal}(K/F)}$. Thus, the condition that (127) are isomorphisms is equivalent to the composite $F \mapsto K^{\mathrm{Gal}(K/F)}$ of (125) followed by (126) being identity.

On the other hand, the condition that the maps (128) are isomorphisms needs to be checked only on transitive continuous $\mathrm{Gal}(K/k)$-sets X, i.e., those of the form $\mathrm{Gal}(K/k)/H$, for subets H that contain $\mathrm{Gal}(K/F_0)$ for some finite subextension F_0/k of K/k. For these H, (128) can be verified to be just the map $G/H \to G/H'$, where $H' = Fix_G(K^H)$. Thus, the assertion that the maps (128) are all isomorphisms is equivalent to the assertion that the composite $H \mapsto Fix_G(K^H)$ of (126) followed by (125) is the identity. Thus, we have shown that the 'equivalence of categories' version is equivalent to the classical version of the theorem, which therefore follows. □

26.5. **Infinite Galois correspondence.** By infinite Galois correspondence, we mean a description of *infinite* subextensions E/k of an infinite Galois extension K/k. We have not discussed this yet.

When K/k is infinite Galois, it is usually no longer true that arbitrary subgroups $H \subset G := \mathrm{Gal}(K/k)$ are fixers H_E of intermediate subfields $k \subset E \subset F$. Galois theory, in this abstract general setting, doesn't give an explicit answer for which H arise as H_E, but says that a topology determined by finite subextensions F/k of E/k can be used to package the answer. Before going ahead, let us also remark that while the infinite Galois correspondence also has an 'equivalence of categories' version, right now we will study the classical version for concreteness.

In this subsection, we will typically write E for a possibly infinite intermediate field extension $k \subset E \subset K$, and F for an intermediate field extension $k \subset F \subset K$ with $[F : k]$ finite.

Note that if $k \subset F_1 \subset F_2 \subset K$ are intermediate extensions, so that $\mathrm{Gal}(K/F_2) \subset \mathrm{Gal}(K/F_1)$, we have canonical maps

$$\mathrm{Gal}(K/k) \to \mathrm{Hom}_{k\text{-}Alg}(F_2, K) \cong \mathrm{Gal}(K/k)/\mathrm{Gal}(K/F_2) \to \mathrm{Gal}(K/k)/\mathrm{Gal}(K/F_1) \cong \mathrm{Hom}_{k\text{-}Alg}(F_1, K).$$

Therefore, the following map of sets makes sense:

$$\mathrm{Gal}(K/k) \to \varprojlim_{\substack{k \subset F \subset K \\ F/k \text{ finite}}} \mathrm{Hom}_{k\text{-}Alg}(F, K). \tag{129}$$

Here, to make sense of the above limit, we used that finite subextensions F/k of K/k, ordered under inclusion, form a directed system (use composita).

If we restrict F to Galois extensions, then any element of $\mathrm{Hom}_{k\text{-}Alg}(F, K)$ has image in $F \subset K$, and hence $\mathrm{Hom}_{k\text{-}Alg}(F, K)$ identifies with $\mathrm{Gal}(F/k)$. Therefore, we also get a map, this time a group homomorphism:

$$\mathrm{Gal}(K/k) \to \varprojlim_{\substack{k \subset F \subset K \\ F/k \text{ finite Galois}}} \mathrm{Gal}(F/k). \tag{130}$$

Lemma 26.20. *Sending E to $H_F = Fix_G(E) = \mathrm{Gal}(K/E)$, and H to K^H, determines a bijection*

(131)

$$\{\textit{Intermediate extensions } k \subset E \subset K\} \to \{_{\textit{each } H_i = \mathrm{Gal}(K/F_i) \textit{ for some } k \subset F_i \subset K,\ [F_i : k] < \infty.}^{\textit{Arbitrary intersections } \bigcap_i H_i,}\}$$

Proof. For an intermediate extension $k \subset F \subset K$ with $[F : k]$ finite, we saw in Lecture 25 (the "easy half", Lemma 25.27) that the inclusion $F \subset K^{H_F}$ is an equality. This gives an analogue of (131),

(132)

$$\{\text{Intermediate extensions } k \subset F \subset K, \text{ with } [F : k] < \infty\} \to \left\{ \begin{smallmatrix} \text{subgroups of the form } H = \mathrm{Gal}(K/F) \\ \text{for some } k \subset F \subset K, \text{ with } [F : k] < \infty \end{smallmatrix} \right\}.$$

(131) is a formal consequence of this, using the following observations:

- Composita correspond to intersections: If $k \subset E \subset K$ is the compositum of $\{k \subset F_i \subset K\}_{i \in I}$, then inside $\mathrm{Gal}(K/k)$ we have an equality $\mathrm{Gal}(K/E) = \bigcap_i \mathrm{Gal}(K/F_i)$.
- The intermediate extensions $k \subset E \subset K$ are precisely the composita of collections of finite intermediate extensions $k \subset F \subset K$.

□

In the above setting, if E/k is Galois if and only if $H_E \subset \mathrm{Gal}(K/k)$ is normal, in which case restriction defines an isomorphism $\mathrm{Gal}(K/k)/\mathrm{Gal}(K/E) \to \mathrm{Gal}(E/k)$ – this was part of the "easy half", Theorem 25.25 from Lecture 25, which covered infinite extensions.

The point of the infinite Galois correspondence seems to be to just systematize the right-hand side of (131): is there a package to describe the intersections $\bigcap_i H_i$ that arise in it?

To this end, we will need to define a topology on $\mathrm{Gal}(K/k)$, for which we will use:

Lemma 26.21. *Let K/k be an arbitrary Galois extension. Then the map* (129) *is a bijection of sets, and the map* (130) *is an isomorphism of groups.*

Proof. Every element of K, being algebraic over k, is contained in a finite subextension F_0/k, which has a finite Galois closure F/k inside K. [82] Thus, giving an element of $\mathrm{Gal}(K/k)$ is equivalent to giving elements $\sigma_F \in \mathrm{Gal}(F/k)$, for each intermediate extension $k \subset F \subset K$ with F/k finite Galois, such that the 'inverse limit compatibility' is satisfied: if $k \subset F_1 \subset F_2 \subset K$ finite Galois, then $\sigma_{F_2}|_{F_1} = \sigma_{F_1}$. This gives that (130) is an isomorphism of groups. Analogous considerations give that (129) is a bijection of sets, finishing the proof of the lemma. □

[82] If F_0 is obtained by adjoining to k roots of polynomials $f_1, \ldots, f_n$ in $k[x]$, then the subfield of K generated by all the roots of the f_i in K is Galois, and independent of the choice of $f_1, \ldots, f_n$ in K.

Recall inverse limits associated to inverse systems in Set and Top. For Set, if $f_{ij} : X_j \to X_i$ are the transition morphisms in an inverse system with $\{X_i\}_i$ as the underlying sets, we have an identification of sets that we already used above:

$$\varprojlim_i X_i = \Big\{(x_i)_i \in \prod_i X_i \mid f_{ij}(x_j) = x_i \,\forall\, i \leqslant j\Big\}.$$

Moreover, the same description is valid in the category Top, where the right hand side acquires the subspace topology from the product topology on $\prod_i X_i$. It is also the weakest topology with respect to which each of the projection morphisms $\mathrm{pr}_i : X \to X_i$ is continuous.

We will not prove the following theorem. We will not use it in an essential way either, but it is good to keep it in mind.

Theorem 26.22. *(i) Given a topological space X, the following are equivalent:*

(a) X is an inverse limit of finite discrete topological spaces: we have a homeomorphism

$$X \to \varprojlim_i X_i,$$

where the right-hand side is an inverse limit in the category Top, and each X_i is a finite set given the discrete topology.

(b) X is compact, Hausdorff and totally disconnected.

(ii) Given a topological group G, the following are equivalent:

(a) G is an inverse limit of finite discrete topological groups: we have a homeomorphic isomorphism of groups

$$G \to \varprojlim_i G_i$$

where the right-hand side is an inverse limit in the category $TopGrp$, and each G_i is a finite group given the discrete topology.

(b) G is compact, Hausdorff and totally disconnected.

Definition 26.23. (i) An inverse limit $\varprojlim_i X_i$ as above, with each X_i a finite set, is called a profinite set. Note that each profinite set X, equipped with the bijection $X \to \varprojlim_i X_i$, has a topology acquired from and depending on the bijection.

(ii) Similarly, we define a profinite group, and give profinite groups a topology depending on a realization $G = \varprojlim_i G_i$.

Here is a series of exercises to study profinite topologies:

Exercise 26.24. Suppose $X = \varprojlim_i X_i$, with each X_i finite, and give X the inverse limit topology. Write $(I, \leqslant)$ for the directed set used in the above inverse limits; thus, if I is made into a category the obvious way, the inverse system is a functor $I^{op} \to Top$.

(i) A sub-directed-set I_0 of I is defined in the obvious way (restrict the order relations). We say that $I_0 \subset I$ is cofinal if for all $i \in I$, there exists $i_0 \in I_0$ such that $i \leqslant i_0$. Show that the projection map

$$\varprojlim_{i \in I} X_i \to \varprojlim_{i \in I_0} X_i$$

is an isomorphism of topological spaces. Note that this does not require the X_i to be finite.

For the remaining problems below, fix a cofinal I_0 as above, and make sure to note that those assertions are independent of I_0.

(ii) Note that a subbasis for the topology on X is given by the various $\mathrm{pr}_i^{-1}(x_i)$, with $i \in I$ and $x_i \in X_i$. However, show that this subbasis is in fact a basis, and that we again get a basis if we consider only the $\mathrm{pr}_i^{-1}(x_i)$ with $i \in I_0$ and $x_i \in X_i$.

For this reason, when we talk of a profinite set X, we will not carry around a presentation of X as $\varprojlim X_i$, but only remember the topology on X. A similar comment will apply to profinite groups.

(iii) Now assume that each X_i is a group G_i, and that $G \to \varprojlim_i G_i$ is an inverse limit of the groups G_i. For each i, show that $H_i := \ker(G \to G_i)$ is an open normal subgroup of G. Further, show that the $\{H_i\}$ form a basis of neighborhoods of the identity. Conclude that a topological group G is profinite if and only if $1 \in G$ has a basis of open neighborhoods $\{H_i\}_{i \in I}$, with each $H_i \subset G$ an open normal subgroup.
Note: If G is any topological group and $H \subset G$ is an open subgroup, then H is automatically closed; this is because H is the complement of the non-identity cosets of H in G, each of which is open. In contrast, a closed subgroup H of a topological group G need not be open, unless H is of finite index in G.

(iv) Conclude that if G is a compact, Hausdorff and totally disconnected group, then $1 \in G$ has a basis of neighborhoods consisting of finite index normal subgroups, which are open and closed in G.
Note: You may use Theorem 26.22 to prove this, but note that the profinite group $\mathrm{Gal}(K/k)$ that is of concern to us is explicitly given such a basis (as you can see from the above exercises), so we will not really use this exercise.

(v) If $G = \varprojlim_i G_i$ is a profinite group, $H \subset G$ is a subgroup, and $\Xi = \{H'\}_{H'}$ is a basis of neighborhoods of the identity in G consisting of open normal subgroups, show that the closure $\bar{H}$ of H is given by

$$\bar{H} = \bigcap_{H' \in \Xi} H \cdot H',$$

where we note that each $H \cdot H' \subset G$ is an open normal subgroup. Conclude that the following are equivalent:

(a) H is closed in G.

(b) H can be written as $\bigcap_i H_i$, where each $H_i \subset G$ is a finite index open normal subgroup of G.

Definition 26.25. Let K/k be an infinite (i.e., possibly infinite) Galois extension. Either of (129) and (130) defines the same topology on $G := \mathrm{Gal}(K/k)$ by Exercise 26.24(i). We equip $G = \mathrm{Gal}(K/k)$ with, making it into a profinite topological group. This topology on $\mathrm{Gal}(K/k)$ is called the Krull topology.

Theorem 26.26. *Let K/k be an arbitrary Galois extension, and set $G = \mathrm{Gal}(K/k)$. Sending F to $H_F = Fix_G(F)$, and H to K^H, induces the following bijections:*

(i)

(133)

$$\{\textit{Intermediate extensions } k \subset F \subset K \textit{ with } [F:k] \textit{ finite}\} \to \{\textit{Open subgroups of } \mathrm{Gal}(K/k)\}.$$

(ii)

(134)

$$\{\textit{Intermediate extensions } k \subset F \subset K \textit{ with } [F:k] \textit{ finite Galois}\} \to \{\textit{Open normal subgroups of } \mathrm{Gal}(K/k)\}$$

(iii)

(135) $$\{\textit{Intermediate extensions } k \subset E \subset K\} \to \{\textit{Closed subgroups of } \mathrm{Gal}(K/k)\}.$$

(iv)

(136)

$$\{\textit{Intermediate extensions } k \subset E \subset K \textit{ with } E/k \textit{ Galois}\} \to \{\textit{Closed normal subgroups of } \mathrm{Gal}(K/k)\}.$$

For each intermediate extension $k \subset E \subset K$ with E/k Galois, restriction from K to E induces an isomorphism $\mathrm{Gal}(K/k)/\mathrm{Gal}(K/E) \to \mathrm{Gal}(E/k)$.

Proof. Recall that the 'easy half' of Galois theory was proved for extensions that were possibly infinite: thus, we have $K^{H_E} = E$ for any subextension E/k of K/k. This half also covered the assertion in (iv) regarding $\mathrm{Gal}(E/k)$. Therefore, in each assertion, it suffices to show that the image of $E \mapsto H_E$ is as described.

Moreover, (ii) and (iv) follow respectively from (i) and (iii), using the "easy half" of Galois theory (Lemma 25.27 from Lecture 25), so it is enough to prove (i) and (iii).

For (i), by the bijection (132), which we observed in the proof of Lemma 26.20, it suffices to show that open subgroups of $\mathrm{Gal}(K/k)$ are precisely the $\mathrm{Gal}(K/F)$, as F varies over intermediate extensions $k \subset F \subset K$ with F/k finite. It is immediate that each such $\mathrm{Gal}(K/F)$ is an open subgroup of $\mathrm{Gal}(K/k)$, and it suffices to show that each open subgroup $U \subset \mathrm{Gal}(K/k)$ is of the form $\mathrm{Gal}(K/F)$ for some such F. By Exercise 26.24(ii), U contains a subgroup of the form $\mathrm{Gal}(K/F_0)$, with F_0/k finite. Replacing F_0 with its Galois closure in K, we may and do assume that F_0/k is finite Galois.

Let $\bar{U}$ be the image of U in $\mathrm{Gal}(K/k)/\mathrm{Gal}(K/F_0) = \mathrm{Gal}(F_0/k)$. By the finite Galois correspondence, we can write $\bar{U} = \mathrm{Gal}(F_0/F)$, for some finite subextension F/k of F_0/k. $U \subset \mathrm{Gal}(K/k)$ is the preimage of $\bar{U} = \mathrm{Gal}(F_0/F) \subset \mathrm{Gal}(F_0/k)$, and hence equals $\mathrm{Gal}(K/F)$. This gives (i).

Given Lemma 26.20, specifically (131), together with Exercise 26.24(v), (iii) follows from (i), finishing the proof of the theorem. □

Exercise 26.27. Let K/k be a Galois extension.

(i) Recall from Definition 26.1 what it means for a action of $\mathrm{Gal}(K/k)$ on a K-vector space V (by k-vector space automorphisms) to be continuous. Show that the continuity of this action is equivalent to the continuity of the action map $\mathrm{Gal}(K/k) \times V \to V$, where $\mathrm{Gal}(K/k)$ is given the Krull topology defined above, and V is given the discrete topology.
(ii) Prove an analogous result about what it means for an action of $\mathrm{Gal}(K/k)$ on a finite set to be continuous (see Definition 26.10(ii)).

Exercise 26.28. Read up about, or work out, an 'equivalence of categories' version of the infinite Galois correspondence, involving profinite (instead of finite) sets with a continuous $\mathrm{Gal}(K/k)$-action. One source for this is the book of Janelidze and Borceux. I would have liked to work it out here, but I don't have time, at least any more than Fermat had margin. It would be (modulo trade-offs/opportunity cost) helpful to then deduce the 'infinite subextension' case from the 'equivalence of categories' version.

26.6. **Appendix – a bit on Galois categories, without proofs.** The material of this subsection is recommended but optional. But even more than reading the following, I encourage you to work through (at least if you have enough time and interest) the material on Galois categories in Lenstra's notes at:
https://websites.math.leidenuniv.nl/algebra/GSchemes.pdf ,
and translate it back to the case of field extensions. Galois categories unify the formal aspects of the Galois theory field extensions and covering spaces. I wanted to write in some detail about them, but don't have the time to do so. However, I will set up notation and state the 'main theorem of Galois theory' in that setting. The reference is the article of Lenstra mentioned above, which itself may have been taken from or inspired by analogous material in SGA1.

The following notion of Galois category axiomatizes some conditions that turn out to be sufficient for a category $\mathcal{C}$ to be equivalent to a category of the form $(\pi\text{-}FinSet)_{cts}$, for a profinite group π. As you read the conditions in the definition below, consider doing the easy verification that these conditions are all satisfied by each $(\pi\text{-}FinSet)_{cts}$, and hence necessary for $\mathcal{C}$ to be equivalent to some $(\pi\text{-}FinSet)_{cts}$.

Definition 26.29. Let $\mathcal{C}$ be a category, [83] and $F : \mathcal{C} \to FinSet$ a functor. We say that $\mathcal{C}$ is a Galois category, with F as a fundamental functor, if the following conditions of Grothendieck are satisfied:

(G1) $\mathcal{C}$ is closed under finite limits (i.e., it has a final object and is closed under fiber products).
(G2) $\mathcal{C}$ is closed under finite coproducts (and in particular has an initial object), and for every finite group H of automorphisms of an object X of $\mathcal{C}$, a 'categorical' quotient X/H exists – the existence of this quotient means, by definition, that (noting that

[83] I will assume that all categories are "essentially small".

H acts on the set $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ for any $Y \in \mathrm{Ob}\,\mathcal{C}$), the functor $(h^X)^H : \mathcal{C}^{op} \to Set$ given by $Y \mapsto \mathrm{Hom}_{\mathcal{C}}(X,Y)^H$ is representable. (Note that this latter 'quotient' condition is saying that $\mathcal{C}$ is closed under a particular kind of coequalizers).

(G3) Any morphism in $\mathcal{C}$ is a composite of an epimorphism followed by a monomorphism, and any monomorphism $X \to Y$ in $\mathcal{C}$ factors in the form $(Y' \hookrightarrow Y' \sqcup Y'') \circ (X \to Y')$, where $X \to Y'$ is an isomorphism.

(G4) F preserves finite limits (i.e., it takes a final object in $\mathcal{C}$ to one in π-$FinSet$, and commutes with fiber products).

(G5) F commutes with finite coproducts, sends epimorphisms to epimorphisms, and commutes with passages $X \to X/H$ to quotients by the action of a finite group H of automorphisms (i.e., the obvious map $F(X)/H \to F(X/H)$ is an isomorphism).

(G6) F is a conservative functor, i.e., $F(u)$ is an isomorphism if and only if u is an isomorphism.

Example 26.30. (i) One example is the category $\mathcal{C}$ of finite covering spaces of a topological space X with a chosen base-point x, with $F(Y \to X)$ being the preimage of x in Y, or equivalently the set of maps $\tilde{X} \to Y$ lifting the universal covering $\tilde{X} \to Y$.

(ii) Another example is the category $\mathcal{C} = (f\acute{e}t_k)^{op}$ opposite to the category $f\acute{e}t_k$ of finite separable algebras over a field k, and where $F(K/k)$ equals $\mathrm{Hom}_{k\text{-}Alg}(K, k^s)$ for a chosen separable closure $k \hookrightarrow k^s$ of k. Note that in this case an initial object of $\mathcal{C}$, whose existence is required by the axiom (G2), corresponds to a final object of $f\acute{e}t_k$, which is the 0-ring viewed as a k-algebra!

Remark 26.31. The axioms in Definition 26.29 are not trivial to verify in the cases mentioned in Example 26.30, and in the case of $(f\acute{e}t_k)^{op}$ in fact include the substantial results from classical Galois theory. For instance, the condition $F(X/H) \cong F(X)/H$ from (G5), in the situation of $\mathcal{C} = (f\acute{e}t)_k^{op}$, implies the following. If $X \in \mathrm{Ob}\,\mathcal{C}$ corresponds to a finite Galois field extension K/k in the opposite category $f\acute{e}t_k$, it is easy to see that X/H corresponds to for the finite separable k-algebra K^H, and the condition in (G5) that $F(X)/H \to F(X/H)$ is an isomorphism implies that the restriction map $\mathrm{Hom}_{k\text{-}Alg}(K, k^s)/H \to \mathrm{Hom}_{k\text{-}Alg}(K^H, k^s)$ is an isomorphism, or equivalently, that $\mathrm{Gal}(K/K^H) = \mathrm{Gal}(K/k)/\bar{H}$, where $\bar{H} \subset \mathrm{Gal}(K/k)$ is the image of H under the restriction map $\mathrm{Gal}(k^s/k) \to \mathrm{Gal}(K/k)$. Thus, this says that the inclusion $\bar{H} \hookrightarrow Fix_{\mathrm{Gal}(K/k)}(K^{\bar{H}})$ is an equality, and is hence the 'latter half' of finite Galois theory.

I think this means that Galois theory is not formal, but the formalism of Galois categories is still useful for us since it helps us give a concrete shape to the analogy between the Galois theories for field extensions and covering spaces, and probably also helps us understand field extensions better.

To prepare for the statement of the main theorem, Lenstra actually considers a second group, the group $\mathrm{Aut}(F)$ of automorphisms of the functor F (this is somewhat analogous to how one realizes suitable abelian categories $\mathcal{A}$ as modules over $\mathrm{Aut}_{\mathcal{A}}(P)^{op}$, where $P \in \mathrm{Ob}\,\mathcal{A}$

is a projective generator: though F itself is not representable in general, it is kind of 'pro-representable'). He then notices that $\mathrm{Aut}(F)$ has an obvious structure of a profinite group (use that F is valued in finite sets), and that F can also be viewed as a functor

$$H : \mathcal{C} \rightsquigarrow (\mathrm{Aut}(F)\text{-}FinSet)_{cts}.$$

In this setting, the main theorem of Galois theory can be stated as (I am basically copying from Theorem 3.5 of Lenstra's notes mentioned above):

Theorem 26.32. *Let $\mathcal{C}$ be an essentially small Galois category with fundamental functor F.*

(i) The functor $H : \mathcal{C} \rightsquigarrow (\mathrm{Aut}(F)\text{-}FinSet)_{cts}$ is an equivalence of categories.

(ii) For any profinite group π, and any equivalence $\mathcal{C} \rightsquigarrow (\pi\text{-}FinSet)_{cts}$ of categories such that the composition

$$\mathcal{C} \rightsquigarrow (\pi\text{-}FinSet)_{cts} \overset{\text{Forget}}{\to} FinSet$$

is naturally isomorphic to F, π is canonically isomorphic to $\mathrm{Aut}(F)$.

(iii) Any two fundamental functors on $\mathcal{C}$ are naturally isomorphic.

(iv) For any profinite group π such that $\mathcal{C}$ and $(\pi\text{-}FinSet)_{cts}$ are equivalent, there is an isomorphism $\pi \to \mathrm{Aut}(F)$ that may not be unique, but is unique up to an inner automorphism of $\mathrm{Aut}(F)$.

I would have liked to describe more of the proof, but as I said I don't have time to do so (or even to work through that proof myself in more detail). It seems to me that reading the proof in Lenstra's notes and comparing it with the usual Galois theory for fields, can help us understand the latter better, and also give some general practice in category theory. For instance, when $\mathcal{C} = (f\acute{e}t_k)^{op}$, you can check that the definition of a connected object in Lenstra's notes corresponds to a finite separable k-algebra that is a field extension. He defines a Galois object of $\mathcal{C}$ to be an object A such that $A/\mathrm{Aut}_{\mathcal{C}}(A)$ is the final object 1 of $\mathcal{C}$; for $\mathcal{C} = (f\acute{e}t_k)^{op}$, this corresponds to the initial object of $f\acute{e}t_k$, namely, the k-algebra k. It is interesting how the analogue of constructing a Galois extension containing a given separable extension, is carried out in this setting, and then how it translates back to the case of $(f\acute{e}t_k)^{op}$: one can no longer take splitting fields associated to polynomials, so one takes a $\#\,\mathrm{Hom}_k(K, k^s)$-fold tensor product of K with itself over k, and looks at a 'connected component' inside it; see Section 3.14 of Lenstra's notes for more details. This is probably a good exercise in field theory, one that could have been at least an optional problem in one of the homework sets.

A different treatment of this material can be found in the stacks project:
https://stacks.math.columbia.edu/tag/0BMQ .

However, the stacks project treatment makes stronger assumptions: e.g., rather than just require commutativity with passage to finite quotients, which was one of the inputs from Galois theory, the stacks project page requires F to be exact, and hence in particular to commute with all finite colimits (which very nontrivially includes coequalizers associated to the action of a finite group).

27. Lecture 27 – Additional topics related to Field and Galois theory (crude)

27.1. **Normal basis theorem.** Let K/k be a finite Galois field extension. Then $G := \mathrm{Gal}(K/k)$ acts on K, and hence K is a left module over $k[G]$.

Theorem 27.1 (Normal basis theorem)**.** *The left $k[G]$-module K (i.e., the k-linear representation of G on the k-vector space K via the Galois action) is isomorphic to $k[G]$ with its right regular action of G. In other words, there exists $\alpha \in K$ such that the Galois orbit $\{\sigma(\alpha) \mid \sigma \in \mathrm{Gal}(K/k)\}$ of α is a basis for K as a vector space over k.*

First we recall a computation from Lecture 26: the base-change of K to K, i.e., $K \otimes_k K$ as a $K[G]$-module, is isomorphic to $K[G]$.

Lemma 27.2. *$K \otimes_k K$, where G acts via the first copy of K, and thought of as a K-vector space via the second copy of K, is isomorphic to $K[G]$ with its right regular representation.*

Proof. This was proved in Lecture 26 (see Proposition 26.15 and Remark 26.16), but let us briefly recall the computation. We have an isomorphism of K-vector spaces

$$K \otimes_k K \cong \mathrm{Maps}(G, K), \qquad \text{such that} \qquad a \otimes b \mapsto (\sigma \mapsto \sigma(a)b).$$

Under this action, the action of $\tau \in G = \mathrm{Gal}(K/k)$ on the first copy of K sends $(\sigma \mapsto \sigma(a)b)$ to $(\sigma \mapsto (\sigma\tau(a))b)$, which is the right regular action on $\mathrm{Maps}(G, K)$. But then $\mathrm{Maps}(G, K)$ with its right regular action is precisely $K[G]$ with its right regular action. □

Lemma 27.3. *Let V, W be representations of a k-algebra A that are finite dimensional as k-vector spaces, and let K/k be a field extension. Suppose $V \otimes_k K \cong W \otimes_k K$ as $A \otimes_k K$-modules. Then $V \cong W$ as A-modules.*

Proof of Theorem 27.1, assuming Lemma 27.3 for finite K/k. Take $A = k[G]$, $V = K$ and $W = k[G]$, where we give W the right regular representation. Then by Lemma 27.2, $V \otimes_k K \cong K \otimes_k K$ and $W \otimes_k K \cong k[G] \otimes_k K = K[G]$ are isomorphic as modules over $K[G] = k[G] \otimes_k K$. Therefore, by Lemma 27.3, and using that K/k is finite, we conclude that $K \cong k[G]$ as $k[G]$-modules. □

Proof of Lemma 27.3, only sketched when K/k is infinite. This may be due to Deuring. First we assume $[K : k] < \infty$. Write

$$V \cong \bigoplus_i V_i^{\oplus m_i}, \qquad W = \bigoplus_i V_i^{\oplus n_i}$$

as a direct sum of indecomposable A-modules (some of the m_i and the n_i may be zero): we have such a decomposition since V and W are finite dimensional as k-vector spaces, and hence of finite length as A-modules. It is enough to show that $m_i = n_i$ for each i.

Note that, as modules over $A \cong A \otimes \{1\} \subset A \otimes_k K$, we have

$$V \otimes_k K \cong \bigoplus_i V_i^{\oplus m_i \cdot [K:k]}, \qquad W = \bigoplus_i V_i^{\oplus n_i \cdot [K:k]}.$$

Therefore, by the Krull-Schmidt-Remak decomposition, we have $m_i \cdot [K : k] = n_i \cdot [K : k]$ for all i, so that $m_i = n_i$ for all i, as desired.

Now consider the general case; we will sketch how to reduce this to the case where K/k is finite. The isomorphism $\varphi : V \otimes_k K \cong W \otimes_k K$ involves only finitely many elements of K. Thus, it is easy (exercise!) to see that there exists a finitely generated k-subalgebra $B \subset K$ such that $V \otimes_k B \cong W \otimes_k B$ as $A \otimes_k B$-modules. Let $\mathfrak{m} \subset B$ be any maximal ideal. Then, with $F := B/\mathfrak{m}$, we have:

$$V \otimes_k F \stackrel{\cong}{\to} W \otimes_k F$$

as modules over $A \otimes_k F$.

Thus, it is enough to prove that $[F : k]$ is a finite extension. Since F is a field extension of K which is finitely generated as a k-algebra, this follows from a result known as Hilbert's Nullstellensatz. □

Remark 27.4. (i) If K/k is Galois but not necessarily finite, one still has a normal basis theorem, which says that with $G = \mathrm{Gal}(K/k)$, we have a G-equivariant isomorphism

$$K \to \mathrm{Maps}_{\mathrm{cts}}(G, k),$$

where the right-hand side is the space of continuous maps $G \to k$ with the left or right regular action of G, k is given the discrete topology, and G is given the Krull topology: see Theorem 1 of the paper "A normal basis theorem for infinite Galois extensions" by H. W. Lenstra Jr., Indagationes Mathematicae (Proceedings), Volume 88, Issue 2, 1985, Pages 221-228 (it is a short paper, you can easily read it). Theorem 2 of the paper gives another generalization of the normal basis theorem to infinite Galois extensions.

(ii) For a more standard proof of the normal basis theorem, using linear independence of characters, see
https://kconrad.math.uconn.edu/blurbs/galoistheory/linearchar.pdf
I haven't read the argument in Serge Lang's book carefully, but at least one of the editions was supposed to have an omission.

27.2. **Inseparable extensions.**

Notation 27.5. We will continue to write $[E : k]_s$ for $\#\,\mathrm{Hom}_{k\text{-}Alg}(E, L)$, where L is an algebraically closed field containing k, even when E/k is only algebraic but not finite. But, at least for now, we consider $[E : k]_i = [E : k]/[E : k]_s$ to be only defined when E/k is finite.

Let us slightly generalize our definition of purely inseparable extensions – the following terminology may be nonstandard (and a less than necessary detour).

Definition 27.6. (i) If A_1 is a finite commutative k-algebra and A is a finite commutative A_1-algebra, then we say that A/A_1 is a purely inseparable extension, if

for every algebraically closed field L containing k, the fibers of the restriction map $\mathrm{Hom}_{k\text{-}Alg}(A, L) \to \mathrm{Hom}_{k\text{-}Alg}(A_1, L)$ are all singleton.

Note that, when A_1 is a field, this terminology agrees with our earlier definition, namely, that A/A_1 is purely inseparable if and only if $[A : A_1]_s = 1$.

(ii) An algebraic field extension K/k is said to be purely inseparable if $[K : k]_s = 1$ – i.e., if there is at most one k-algebra homomorphism from K into any given (algebraically closed, or any other) field. Note that when K/k is finite, this definition agrees with (i). For an algebraic field extension K/k, $\alpha \in K$ is said to be purely inseparable over k if $k[\alpha]/k$ is a purely inseparable extension.

Example 27.7. If k is algebraically closed, a finite commutative k-algebra A is purely inseparable over k if and only if A is local.

Recall from Lecture 24 that any finite commutative k-algebra A has a maximal separable k-subalgebra A_0. Let us study A_0 better.

Proposition 27.8. *(i) If A is a finite commutative k-algebra, and F/k is an arbitrary (not necessarily algebraic) field extension, then A/k is purely inseparable if and only if $A \otimes_k F/F$ is.*

(ii) Purely inseparable finite commutative algebras over k are closed under taking quotients (by ideals) and tensor products. Moreover, if $k \subset A_1 \subset A_2 \subset A$ with A/k purely inseparable, then A_2/A_1 is purely inseparable.

(iii) If A is a finite commutative algebra over k which is local, and $\{A_i\}_{i\in I}$ is a family of purely inseparable subalgebras of A, then the subalgebra of A generated by the A_i is purely inseparable.

(iv) If a field extension K/k is a purely inseparable algebraic field extension, and $k \subset F \subset E \subset K$ are intermediate extensions, then E/F is purely inseparable.

(v) If $k \subset E \subset K$ are field extensions with E/k and K/E purely inseparable, then so is K/k.

(vi) If E, F are contained in an algebraic field extension K/k, and if E/k is purely inseparable, so is EF/F. Note that this property, together with (v), implies that if both E/k and F/k are purely inseparable, so is EF/k.

Proof. Unless otherwise stated, L will denote an algebraically closed field containing k.

(i) is immediate from the fact that separable degree is invariant under base-change (Proposition 24.22(i) from Lecture 24).

If a k-algebra A' is a quotient of a purely inseparable k-algebra A, then $\mathrm{Hom}_{k\text{-}Alg}(A', L) \subset \mathrm{Hom}_{k\text{-}Alg}(A, L)$ has cardinality 1. If A', A'' are purely inseparable k-algebras, then $\mathrm{Hom}_{k\text{-}Alg}(A' \otimes_k A'', L) \cong \mathrm{Hom}_{k\text{-}Alg}(A', L) \times \mathrm{Hom}_{k\text{-}Alg}(A'', L)$ is singleton, giving closure under taking tensor products. If $k \subset A_1 \subset A_2 \subset A$ are finite commutative k-algebras, then for any k-algebra homomorphism $A_1 \to L$ with L algebraically closed, we have

$$\mathrm{Hom}_{k\text{-}Alg}(A, L) \supset \mathrm{Hom}_{A_1\text{-}Alg}(A, L) \twoheadrightarrow \mathrm{Hom}_{A_1\text{-}Alg}(A_2, L), \tag{137}$$

where the latter map is surjective, as follows from the analogous result for fields. If A/k is purely inseparable, then $\mathrm{Hom}_{k\text{-}Alg}(A, L)$ is singleton, and hence so is $\mathrm{Hom}_{A_1\text{-}Alg}(A_2, L)$ by (137), so that A_2/A_1 is purely inseparable. This finishes the proof of (ii).

(iii) is proved as in the separable case (see Corollary 24.33(ii) from Lecture 24), using closure under tensor products.

For (iv), follow the proof of the analogous assertion in (ii). (v) is asserting the following: if there is only one k-algebra homomorphism $E \to L$, and if that homomorphism $E \to L$ extends to only one homomorphism $K \to L$, then there is only one k-algebra homomorphism $K \to L$; this is clear. For (vi), note that for suitably large L, $\mathrm{Hom}_{F\text{-}Alg}(EF, L)$ injects into $\mathrm{Hom}_{k\text{-}Alg}(E, L)$. □

Corollary 27.9. *Let K/k be an algebraic field extension. The following are equivalent:*

(i) K/k is purely inseparable, i.e., $[K : k]_s = 1$.
(ii) Each $\alpha \in K$ is purely inseparable over k.
(iii) K is generated over k by a family of elements each of which is purely inseparable over k.

Proof. Easy exercise, using ideas that we have already seen. □

If K/k is a purely inseparable field extension, it is clear that it has no subextension K_0/k which is separable (please do this as an exercise if this is not clear to you). However, the converse doesn't seem clear from what we have seen so far today, and will be proved below.

27.3. **Field extensions and inseparability.**

Lemma 27.10. *Let K/k be a (not necessarily finite) field extension.*

(i) There exists a subextension $k \subset K_0 \subset K$ such that K_0/k is separable, and K/K_0 is purely inseparable.
(ii) Any K_0 as in (i) is the unique maximal separable subextension of K.

Proof. If $\mathrm{char}\, k = 0$, K/k is separable and there is nothing to prove, so let us assume $\mathrm{char}\, k = p > 0$.

First we prove the assertions assuming that K is a finite extension of k. Since $\mathrm{char}\, k = p > 0$, $x \mapsto x^p$ is a field homomorphism of K, and its image K^p is a subfield of K. Similarly, we can talk of K^{p^n} for any $n \geqslant 0$, and of the compositum $K^{p^n}k$ of K^{p^n} and k in K.

We claim that for each n, $K/K^{p^n}k$ is purely inseparable. If $\sigma, \sigma' : K \hookrightarrow L$ are $(K^{p^n}k)$-algebra homomorphisms into an algebraically closed field L containing $K^{p^n}k$, then for all $\alpha \in K$ we have that $\sigma(\alpha^{p^n}) = \sigma'(\alpha^{p^n})$, so $\sigma(\alpha) = \sigma'(\alpha)$, because p^n-th roots are unique in characteristic p (if $x^{p^n} = y^{p^n}$, then $(x - y)^{p^n} = 0$, so $x - y = 0$). Thus, $\mathrm{Hom}_{K^{p^n}k\text{-}Alg}(K, L)$ is singleton, yielding our claim that $K/K^{p^n}k$ is purely inseparable.

Since $[K : k]$ is finite, we have for large enough n that $K^{p^n}k = K^{p^{n+1}}k = \ldots$. Choose such an n, and set $K_0 = K^{p^n}k$. To prove (i), it suffices to prove that K_0/k is separable.

Let us first prove this in the special case where $K = k[\alpha]$ for some α. In this case, $K_0 = k[\alpha^{p^n}]$ for all large enough n. Since $K_0^p k = K_0$, we get that if $\beta = \alpha^{p^n}$ for a suitably large n, then $\beta = g(\beta^p)$ for some $g \in k[x]$. In other words, β is a root of $f \in k[x]$, where $f(x) = x - g(x^p)$ is separable, so that $K_0 = k[\beta]$ is separable over k, as desired.

To prove the lemma for general $K = k[\alpha_1, \ldots, \alpha_r]$ (still assuming that K/k is finite). $K_0 = k[\alpha_1^{p^n}, \ldots, \alpha_r^{p^n}]k$ for large n. By the case where $K = k[\alpha]$, we know that each of $k[\alpha_1^{p^n}], \ldots, k[\alpha_r^{p^n}]$ is separable over k, for large n, so their compositum $K_0 = k[\alpha_1^{p^n}, \ldots, \alpha_r^{p^n}]$ is separable over k as well.

Now let us prove (ii) (still assuming that K/k is finite). K_0, being separable over k, is contained in the maximal separable subextension of k in K, say K_0'. If K_0 is properly contained in K_0', we have

$$1 = [K : K_0]_s = [K : K_0']_s [K_0' : K_0]_s,$$

so $[K_0' : K_0]_s = 1$, so that

$$[K_0' : k]_s = [K_0' : K_0]_s [K_0 : k]_s = [K_0 : k]_s \leqslant [K_0 : k] < [K_0' : k],$$

contradicting that K_0' is separable over k. Thus, we are done when K/k is finite.

Now we consider the case where K/k is infinite. In this case, let $K_0 \subset K$ instead denote the maximal separable subextension of K/k. If $\alpha \in K \backslash K_0$, it is easy to see that $K_0 \cap k[\alpha]$ is the maximal separable subextension of $k[\alpha]/k$, so $k[\alpha]$ is purely inseparable over $K_0 \cap k[\alpha]$ by the already proved finite case, and it follows that $K_0[\alpha]$ is purely inseparable over K_0. Thus, K/K_0 is purely inseparable by Corollary 27.9, giving (i).

Given our choice of K_0, (ii) is immediate in this case. □

Corollary 27.11. *If K/k is a finite field extension and $\operatorname{char} k = p > 0$, then K/k is separable if and only if $K^p k = K$.*

Proof. If $K^p k \subsetneq K$, then we saw in the above proof that $K/K^p k$ is purely inseparable, so K/k is not separable. On the other hand, if $K^p k = K$, then $K^{p^n} k = K$ for all n, so the K_0 constructed in the above proof equals K, so $K/k = K_0/k$ is separable. □

Corollary 27.12. *Assume that $\operatorname{char} k = p > 0$. A singly generated finite algebraic extension $k[\alpha]$ of k is purely inseparable if and only if the monic minimal polynomial of α over k is of the form $x^{p^n} - a$, with $a \in k$.*

Proof. By the proof of Lemma 27.10, $k[\alpha]/k$ is purely inseparable if and only if $(k[\alpha])^{p^n} k = k[\alpha^{p^n}]$ equals k for all large enough n. This is equivalent to α satisfying a polynomial over k of the form $x^{p^n} - a = 0$ for some $a \in k$.

A polynomial $f \in k[x]$ is of the form $x^{p^n} - a$ if and only if in some or equivalently any algebraic closure L of k, f splits into the form $(x - \alpha)^{p^n}$ for some n. Thus, the condition that α satisfies a polynomial of the form $x^{p^n} - a = 0$ is equivalent to the condition that the minimal polynomial of α is of the form $x^{p^n} - a$, yielding the corollary. □

Corollary 27.13. *Let K/k be a finite purely inseparable extension, and assume that* $\operatorname{char} k = p > 0$.

(i) $[K : k]$ *is a power of* p.
(ii) For each $\alpha \in K$, *there exists* $n > 0$ *such that* $\alpha^{p^n} \in k$.

Proof. Immediate from Corollary 27.9 and Corollary 27.12. □

Thus, purely inseparable field extensions are obtained by attaching p-power roots successively.

Exercise 27.14. Let $K = k[\alpha]$, and suppose $\operatorname{char} k = p > 0$. Fix any algebraically closed field L containing $k[\alpha]$. Let $f \in k[x]$ be the minimal polynomial of α. Choose n maximal so that $f(x) = g(x^{p^n})$ for some $g \in k[x]$. Show that the roots of f in L all occur with the same multiplicity equal to $[k[\alpha] : k]_i$, and that the maximal separable subextension of $k[\alpha]/k$ is obtained by adjoining to k a root of g in K. Show also that the distinct roots α_i of f in L and the (necessarily distinct) roots β_i of g in L are in bijection, given by $\beta_i = \alpha_i^{p^n}$. Use this to give an alternate proof of Lemma 27.10.

Part of the corollary below has already been proved, e.g., in Corollary 27.11, but it seems to be a convenient summary.

Corollary 27.15. *Given an algebraic field extension K/k, the following are equivalent:*

PI1. $[K : k]_s = 1$, *i.e., K/k is purely inseparable.*
PI2. *The only separable subextension of K/k is k itself.*
PI3. *Each $\alpha \in K$ is purely inseparable over k.*
PI4. *For all $\alpha \in K$, the minimal monic polynomial of α over k is of the form $x^{p^n} - a = 0$, with some $n \geqslant 0$ and $a \in k$.*
PI5. *K is generated by a family $\{\alpha_i\}_{i \in I}$, each of which is purely inseparable over k.*

Proof. **PI1**, **PI3** and **PI5** are equivalent by Corollary 27.11.

If $[K : k]_s = 1$, then for all subextensions E/k of K/k, we have $[E : k]_s = 1$, so **PI1** implies **PI2**. Lemma 27.10 implies that **PI2** implies **PI3**.

Given **PI3**, Corollary 27.12 gives **PI4**. If α satisfies a polynomial of the form $x^{p^n} - a$, then for each k-algebra embedding of $\sigma : k[\alpha] \to L$ with L algebraically closed, $\sigma(\alpha)^{p^n} = a = \alpha^{p^n}$, so we get $\sigma(\alpha) = \alpha$, so $[k[\alpha] : k]_s = 1$. Thus, **PI4** implies **PI3** as well. All the equivalences have been proved.

□

Example 27.16. It is easy to see that any algebraic extension K/k of fields has a maximal purely inseparable subextension K_1/k, namely the compositum of all the purely inseparable subextensions of K/k in K. However, in non-analogy with Lemma 27.10, K/K_1 may not be separable. For instance, let $k = \mathbb{F}_2(y, z)$, the field of rational functions in two variables y and z over the finite field $\mathbb{F}_2$, and let K/k be a degree 4 extension generated by a root of f, where $f(t) = t^4 + yt^2 + z$. It is immediately verified that f is irreducible over $\mathbb{F}_2(y, z)$. Since $f(t) = g(t^2)$, where $g(t) = t^2 + yt + z$, and since g is irreducible and separable, it is easy to see that $[K : k]_s = [K : k]_i = 2$.

Suppose K_1/k is nontrivial and purely inseparable. Then $[K_1 : k] = 2$, and K/k_1 is separable. The prime factorization of f takes the form $f(x) = (x - \alpha)^2(x - \beta)^2$, so the minimal polynomial of f over K_1 is $(x - \alpha)(x - \beta)$, which is a square-root of f. But for an algebraic closure L of $k = \mathbb{F}_q(y, z)$, f has a square-root in $L[t]$ of the form $f_1(t) = t^2 + \sqrt{y}t + \sqrt{z}$. Since square-roots are unique in characteristic p, we conclude that $\sqrt{y}, \sqrt{z} \in K_1$. But this forces $[K_1 : k] \geqslant 4$, a contradiction.

I saw this example in an article by Joseph Lipman called "Balanced field extensions" that I recommend, but clearly examples of this kind must have been known well before his article. In particular, he proves in the article that the following two properties are equivalent, and each of them is in turn equivalent to a third property that we won't concern ourselves with here:

(i) K/k can be factored as $K/K_1/k$, where K/K_1 is separable and K_1/k is purely inseparable.
(ii) Some separable algebraic extension of K is normal over k.

We will see in Proposition 27.17 below that normal algebraic extensions do have such a factorization.

Proposition 27.17. *Let K/k be a normal algebraic extension, and let $G = \mathrm{Aut}_{k\text{-}Alg}(K)$.*

(i) The unique maximal purely inseparable subextension of K/k is given by K^G, the subfield of K fixed by G.
(ii) K/K^G is Galois.
(iii) If $k \subset K_0 \subset K$ is the maximal separable subextension, then $K = K^G K_0$, and $K_0 \cap K^G = k$.

Proof. Let L be an algebraically closed field containing $K \supset k$. The unique maximal purely inseparable subextension of K/k consists of all the $\alpha \in K$ that are purely inseparable over k, i.e., such that

$$\Xi_\alpha := \{\sigma(\alpha) \mid \sigma \in \mathrm{Hom}_{k\text{-}Alg}(k[\alpha], L)\} = \{\sigma(\alpha) \mid \sigma \in \mathrm{Hom}_{k\text{-}Alg}(K, L)\} = G \cdot \alpha$$

is singleton (use the normality of K and the fact that L is algebraically closed, to get the latter two equalities in the above line), i.e., of all the $\alpha \in K$ fixed by G.

This proves (i), but note that that does not imply (ii). Instead, we have seen from Galois theory (Lemma 25.30 in Lecture 25) that K/K^G is Galois and hence separable, at least in the case of finite G, but one reduces to the case of finite G by considering finite normal subextensions K_1/K^G of K/K^G. This proves (ii).

Since $K_0 \cap K^G$ is both separable and purely inseparable over k, it follows that $K^0 \cap K^G = k$. Since $K/K^G K^0$ is both separable (since K/K^G is separable) and purely inseparable (since K/K^0 is purely inseparable), it follows that $K = K_0 K^G$. □

Definition 27.18. A field k is called perfect if either $\operatorname{char} k = 0$, or if $\operatorname{char} k = p > 0$ and $k = k^p$.

Proposition 27.17 has the following corollary for perfect fields.

Corollary 27.19. *The following are equivalent:*

(i) k is perfect.
(ii) k has no nontrivial purely inseparable extension.
(iii) Every algebraic extension of k is separable.
(iv) Every algebraic extension of k is perfect.

Proof. Note that the given conditions all automatically hold when $\operatorname{char} k = 0$, so let us assume $\operatorname{char} k = p > 0$.

The equivalence (i) ⇔ (ii) needs just the definition, as follows. k has no nontrivial purely inseparable extension if and only if it has no singly generated nontrivial purely inseparable extension. If k is perfect, then it has no singly generated nontrivial algebraic extension $k[\alpha]$, since this would force $\alpha^{p^n} \in k$ for some n, and hence $\alpha \in k$ by perfectness and the fact that p^n-th roots of unity are unique in characteristic p. Conversely if k is not perfect, then any $a \in k \backslash k^p$ gives a nontrivial purely inseparable extension $k[\alpha]$ where α is the unique p-th root of a.

It is for (ii) ⇒ (iii) that one uses Proposition 27.17. Namely, K is contained in some normal extension K'/k (e.g., K' is contained in an algebraic closure), which Proposition 27.17 lets us write as the composition of a purely inseparable extension and a separable extension. Hence (ii) gives us that K'/k is separable, and hence so is K/k, giving (iii).

If every algebraic extension K/k is separable, then it also follows that such an extension K/k has no nontrivial purely inseparable extension – otherwise, if K'/K is purely inseparable and nontrivial, it follows that K'/k is an algebraic extension that is not separable, a contradiction. This gives that every algebraic extension K/k satisfies (ii), and hence by the equivalence of (i) and (ii), also (i), giving (iv). Finally, it is immediate that (iv) implies (i). □

27.5. **Maximal separable subalgebra.**

Lemma 27.20. *Let $A = \prod_{i=1}^r A_i$ be a finite commutative k-algebra, with each A_i Artin local. For $1 \leqslant i \leqslant r$, let $\mathfrak{m}_i \subset A_i$ be the unique maximal ideal, and let $K_i = A_i/\mathfrak{m}_i$.*

(i) *The maximal separable subalgebra $A_0 \subset A$ takes the form $A_0 = \prod_{i=1}^r A_{0,i}$, where for each $1 \leqslant i \leqslant r$, $A_{0,i} \subset A_i$ is a maximal separable subalgebra.*
(ii) *For each $1 \leqslant i \leqslant r$, $A_{0,i} \to A_i \to A_i/\mathfrak{m}_i = K_i$ defines a k-algebra isomorphism from $A_{0,i}$ to the maximal separable subalgebra of the k-algebra K_i.*
(iii) *Let $(A/\operatorname{rad}(A))_0$ denote the maximal separable subalgebra of $A/\operatorname{rad}(A)$. Then the map $A_0 \hookrightarrow A \to A/\operatorname{rad}(A) \supset (A/\operatorname{rad}(A))_0$ defines an isomorphism from A_0 to $(A/\operatorname{rad}(A))_0$.*

In particular, $A_0 \cong \prod_{i=1}^r K_{0,i}$, where for $1 \leqslant i \leqslant r$, $K_{0,i} \subset K_i$ is the maximal separable subextension of K_i/k.

Proof. (iii) and the last assertion are a summary of the combination of (i) and (ii), so let us prove the first two assertions.

We let A_0 be the maximal separable k-subalgebra of A, but define, for $1 \leqslant i \leqslant r$, $A_{0,i}$ to be the image of A_0 under $A \to A_i$. Since separable algebras are closed under taking homomorphic images, each $A_{0,i}$ is a separable k-algebra. We now have an inclusion $A_0 \hookrightarrow \prod_{i=1}^r A_{0,i}$, which is an equality since $\prod_{i=1}^r A_{0,i}$, being a product of separable k-algebras, is separable. This gives (i).

Let $K_{0,i} \subset K_i$ be the maximal separable k-subalgebra. Since separability of finite commutative k-algebras is preserved under taking quotients, $A_{0,i} \to K_i$ has image contained in $K_{0,i}$. Moreover, each $A_{0,i}$, being reduced and hence a product of fields, and contained in the Artin local ring A_i, is a field. Therefore, $A_{0,i} \to K_{0,i}$ is injective; view it as an inclusion.

Suppose it is not surjective. Let $\bar{\alpha} \in K_{0,i} \backslash A_{0,i}$. The minimal polynomial of $\bar{\alpha}$ is some separable polynomial $f \in k[x]$. By Hensel's lemma (Theorem 17.23 from Lecture 17), applied to the homomorphism $A_i \to K_i$ of rings with nilpotent kernel $\mathfrak{m}_i$, the root $\bar{\alpha} \in K_{0,i}$ of $f \in k[x] \subset K_{0,i}[x]$ lifts to a root $\alpha \in A_i$ of $f \in k[x] \subset A_{0,i}[x]$. Since f is a separable irreducible polynomial, the subring $A_{0,i}[\alpha] \subset A_i$ generated by α over $A_{0,i}$ is a field, and is separable over $A_{0,i}$ and hence over k. This contradicts that $A_{0,i}$ is a maximal separable subalgebra of A_i. This gives the surjectivity of $A_{0,i} \to K_{0,i}$, which is the assertion of (ii). □

Example 27.21. Note that the algebra $A = \mathbb{R}[x]/(x^2+1)^2$ over $k = \mathbb{R}$ is local, since the ideal $(x^2+1) \subset A$ is nilpotent and maximal (maximal since going modulo it gives us a field $K = \mathbb{C}$). It follows from Lemma 27.20 above that the maximal separable subalgebra of A is isomorphic to $\mathbb{C}$ (of course, this follows from Hensel's lemma, which was used in the proof of Lemma 27.20). If you try to find a square-root of -1 in A directly, it can be seen not to be all that simple a polynomial.

Exercise 27.22. (This problem gives a *somewhat* intrinsic version of the Jordan decomposition of linear operators – in the context of finite k-algebras, 'independently of a representation of the algebra on a vector space realizing the algebra as an algebra of linear operators'). Let V be a finite dimensional vector space over a field k, and let $T \in \operatorname{End}_k(V)$. Recall that T is said to be semisimple if it is diagonalizable over an algebraic closure $k \hookrightarrow \bar{k}$ of k, i.e., viewed as an element of $\operatorname{End}_{\bar{k}}(V \otimes_k \bar{k})$, T is diagonalizable.

(i) Show that T is semisimple if and only if the finite commutative k-algebra $k[T] \subset \mathrm{End}_k(V)$ is a separable k-algebra.
Hint: This is very easy: you can base-change to an algebraically closed field (why?), and there you know how separable algebras look like.

(ii) Let $T \in \mathrm{End}_k(V)$ be arbitrary (not necessarily semisimple). A Jordan decomposition of T is a decomposition $T = T_s + T_n$, where $T_s, T_n \in \mathrm{End}_k(V)$ are endomorphisms with T_s semisimple, T_n nilpotent, and such that T_s and T_n commute with each other (and hence with T as well). By/as in Problem 2 of HW 1, such T_s and T_n exist if k is algebraically closed, and for general k, T_s and T_n are uniquely determined if they exist (since uniqueness may be checked after base-changing to an algebraic closure).

Show that the following are equivalent for $T \in \mathrm{End}_k(V)$:

(a) T has a Jordan decomposition $T = T_s + T_n$ in $\mathrm{End}_k(V)$.

(b) The image of $T \in A := k[T]$ in $A/(\mathrm{rad}\, A)$ belongs to the maximal separable k-subalgebra $(A/(\mathrm{rad}\, A))_0$ of $A/(\mathrm{rad}\, A)$.

(iii) (This is sort of part of, and hence also a hint for, (ii), so it is possible that you may want to prove this simultaneously while proving (ii)). Show that when a Jordan decomposition $T = T_s + T_n$ of $T \in \mathrm{End}_k(V)$ exists, T_s is simply the unique element of the maximal separable k-subalgebra $A_0 \subset A$ that has the same image as T in $(A/(\mathrm{rad}\, A))_0$.
Hint for (ii) and (iii): By Lemma 27.20, $A_0 \to (A/(\mathrm{rad}\, A))_0$ is surjective, where A_0 is the maximal separable k-subalgebra of A.
Note:

(a) Recall from HW 1 Problem 2 that the Jordan decompositions there realized T_s and T_n as polynomials in T. That is sort of baked in to our situation, since T_s and T_n are already in the algebra $k[T]$.

(b) All the above go through to give a 'multiplicative Jordan decomposition' $T = T_s T_u$ of any $T \in \mathrm{GL}_k(V)$ under similar hypotheses (i.e., when the equivalent conditions of (ii) hold), where $T_s \in k[T]$ is semisimple and $T_u \in k[T] \subset \mathrm{End}_k(V)$ is unipotent. Automatically, T_s and T_u commute in $k[T]$.

(c) Make sense of the following in the light of your work on the above exercises: thus, additive and multiplicative Jordan decompositions can be defined on any finite commutative k-algebra A (without any other vector space in sight), with the property that for any finite dimensional k-vector space V and any k-algebra homomorphism $\varphi : A \to \mathrm{End}_k(V)$, $x \in A$ (if we are interested in an additive Jordan decomposition) or $x \in A^\times$ (if we are interested in a multiplicative Jordan decomposition) has a Jordan decomposition if and only if $\varphi(x) \in \mathrm{End}_k(V)$ has, in which case φ transports the Jordan decomposition of x into that of $\varphi(x)$.
Here, note that 'nilpotent', 'unipotent' and 'semisimple' have intrinsic meanings in a k-algebra A independently of any embedding $A \hookrightarrow \mathrm{End}_k(V)$: a nilpotent element of A is just one that is nilpotent as a ring element, and a unipotent element of A is just an element u of A such that $u - 1$ is nilpotent.

The above exercises give an intrinsic meaning to 'semisimple' too: namely, one that belongs to the maximal separable k-subalgebra $A_0 \subset A$, or equivalently to some separable k-subalgebra of A.

(iv) Application: If $g \in GL_n(k)$ is a semisimple element, show that its centralizer in $GL_n(k)$ is isomorphic to a group of the form $\prod_{i=1}^r GL_{m_i}(K_i)$, where each K_i/k is a finite separable extension.
Hint: This is easier than it seems. $V = k^n$ is a module over $k[x]$, where x acts via g, or equivalently through $k[x] \to A := k[g] \subset \mathrm{End}_k(V)$, and the centralizer of g is just $\mathrm{Aut}_A(V) \subset \mathrm{End}_A(V)$. A itself is a product of fields since A/k is a separable algebra.
Note: Note that such a result would be harder to see if one were working with matrices.

(v) (ii) shows that if k is perfect, then every $T \in \mathrm{End}_k(V)$ has a Jordan decomposition. Deduce it independently, from HW 1 problem 2 (including its optional/extra assertions), using the existence and uniqueness of the Jordan decomposition over the algebraic closure $\bar{k}$, and taking $\mathrm{Gal}(\bar{k}/k)$-invariants – note that $\mathrm{Gal}(\bar{k}/k)$ makes sense, because since k is perfect, $\bar{k}$ is separable over k.

We state the following assertion with only an idea of proof; it is not hard, but we don't have time to prove it:

Proposition 27.23. *If A is a finite commutative k-algebra and $A_0 \subset A$ is its maximal separable k-subalgebra, then for any field F, $A_0 \otimes_k F$ is the maximal separable F-subalgebra of $A \otimes_k F$.*

Idea of the proof. Roughly, it is clear that $A_0 \otimes_k F \subset A \otimes_k F$ is a separable F-subalgebra; the point is to show that there is no bigger separable subalgebra in $A \otimes_k F$. For this, we may assume F to be as large as we wish. The general case follows from three cases: the first is when F is a separable closure of k, the second is when k is separably closed and F is an algebraic closure of k, and the third is when both k and F are algebraically closed. The first case is taken care of by Galois descent. The second is proved by noting that passing to a purely inseparable extension cannot introduce new idempotents (because if the characteristic is $p > 0$ and F/k is purely inseparable, then for any idempotent $e \in A \otimes_k F$, $e = e^{p^n} \in A$ for large enough n). The third is proved by noting that for an algebraically closed field k, every finite commutative k-algebra is a product of local k-algebras with residue field k, and this condition is preserved by base-change. □

Remark 27.24. The above lemma generalizes to the case where A is only commutative and finitely generated, and not necessarily finite. The maximal separable subaglebra of A then defines what is known as the 'π_0' of A from the perspective of scheme theory.

Lemma 27.25. *Let A be a finite commutative k-algebra, and let $A_0 \subset A$ a maximal separable k-subalgebra. Then A/A_0 is purely inseparable in the sense of Definition 27.6.*

Proof. We have $A = \prod_{i=1}^r A_i$ and $A_0 = \prod_{i=1}^r A_{0,i}$, with each A_i Artin local, and each $A_{0,i} \subset A_i$ a maximal separable subalgebra.

Let L be an algebraically closed field. Each homomorphism $A \to L$ factors through the projection $A \to A_i$ for a unique i, and it thus suffices to show that $\mathrm{Hom}_{k\text{-}Alg}(A_i, L) \to \mathrm{Hom}_{k\text{-}Alg}(A_{0,i}, L)$ is injective. Each such homomorphism factors through the $\mathrm{rad}(A_i) \supset \mathrm{rad}(A_{0,i}) = 0$ (recall that $\mathrm{rad}(A_i)$ is just the nilradical of A_i), so by Lemma 27.20, according to which $A_{0,i}$ surjects onto the maximal separable k-subalgebra $K_{0,i}$ of K_i, it remains to show that $\mathrm{Hom}_{k\text{-}Alg}(K_i, L) \to \mathrm{Hom}_{k\text{-}Alg}(K_{0,i}, L)$ is injective.

Thus, we are reduced to the case where A/k is a field extension, which is treated in Lemma 27.10. □

27.6. **Norm and trace.**

Definition 27.26. S/R ring extension, S finite free over R. Then $N_{S/R} : S \to R$ is the multiplicative monoid homomorphism $s \mapsto \det(m_s)$, where $(m_s : S \to S) \in \mathrm{End}_R(S)$ is multiplication by S. Recall that we had defined $\mathrm{tr}_{S/R} : S \to R$ to be the additive group homomorphism $s \mapsto \mathrm{tr}(m_s)$.

Since the following is nontrivial, and because it seems subtle enough not to have 'canonical' proofs, let us give it the status of a theorem.

Theorem 27.27. *(i) Let $R \to S \to T$ be ring homomorphisms such that T is finite free over S and S is finite free over R. Then $\mathrm{tr}_{T/R} = \mathrm{tr}_{S/R} \circ \mathrm{tr}_{T/S}$, and $N_{T/R} = N_{S/R} \circ N_{T/S}$.*

(ii) Let $R \to S$ be a ring homomorphism such that S is finite free over R, and let V be a finite free S-module. If $T \in \mathrm{End}_S(V) \subset \mathrm{End}_R(V)$, then $\mathrm{tr}_R(T) = \mathrm{tr}_{S/R}(T) \circ \mathrm{tr}_S(T)$, and $\det_R(T) = N_{S/R}(\det_S(T))$.

An ugly proof, assuming that R is a field. First, some remarks regarding the general case. The assertions involving the trace are easy, and will be skipped. We refer to John Sylvester's paper "Determinants of block matrices":

https://hal.science/hal-01509379/document ,

for a general proof of the assertions regarding the norm: it is an easy, elementary and elegant proof, one that nevertheless uses matrices. I don't reproduce it here because I don't want to write out those matrices. But the rough idea is as follows: the assertions regarding the norm are easy to verify for matrices that are products of block upper or lower triangular matrices, and general matrices can be obtained from these, if certain elements can be guaranteed to be nonzerodivisors, which one ensures by adjoining variables.

I will write out a proof of the transitivity of norm assuming that $R = k$ is a field. But what I have written below has turned out to be ugly, so I am not sure if you will want to read it, but I hope you read up a proof from elsewhere.

(ii) implies (i), so it is enough to prove (ii). Replacing $R = k$ and S by $R \otimes_k \bar{k} = \bar{k}$ and $S \otimes_k \bar{k}$, where $k \hookrightarrow \bar{k}$ is an algebraic closure, we assume without loss of generality that k is algebraically closed.

We can write $S = S_1 \times \cdots \times S_n$, where $S_1, \ldots, S_n$ are Artin local k-algebras. Accordingly, we can write $V = V_1 \oplus \cdots \oplus V_n$, where each V_i is a free S_i-module on which S acts through the projection $S \to S_i$. This allows us to write $T = (T_1, \ldots, T_n)$, where $T_i \in \mathrm{End}_{S_i}(V_i)$ for each i. It is easy to see that $\det_S(T) = (\det_{S_1}(T_1), \ldots, \det_{S_n}(T_n)) \in S_1 \times \cdots \times S_n = S$, and $N_{S/k}(\det_S(T)) = \prod_{i=1}^n N_{S_i/k}(\det_{S_i}(T_i)) \in k$. Since $\det_k(T) = \prod_{i=1}^n \det_k(T_i)$ as well, this reduces us to the case where S is Artin local.

We may write $V = \bar{V} \otimes_k S$, for some k-vector space $\bar{V}$. Since S is Artin local, the composite $k \hookrightarrow S \to \bar{S} := S/(\mathrm{rad}\, S)$ is an isomorphism. Therefore, we may also identify $\bar{V}$ with the vector space $V \otimes_S (S/(\mathrm{rad}\, S)) = V \otimes_S k$. This lets us realize $\mathrm{End}_k(\bar{V})$ as a subring of $\mathrm{End}_S(V)$ as well as a quotient ring.

To finish, it is enough to prove:

$$\det_k(T) = \overline{\det_S(T)}^{[S:k]} = N_{S/k}(\det_S(T)), \tag{138}$$

where we write $a \mapsto \bar{a}$ for $S \to \bar{S}$.

Note that the first equality of (138), applied to the case where V is replaced by S and T by multiplication by $\det_S(T)$, implies the second. Therefore, it is enough to prove the first equality of (138).

For this, we claim that, if T_1 and T_2 have the same image in $\mathrm{End}_k(\bar{V})$, then $\det_k(T_1) = \det_k(T_2)$. To see this claim, note that $T_1 - T_2$ sends V to $\mathrm{rad}(S) \cdot V$, so T_1 and T_2 induce the same k-linear maps on the successive quotients of the (finite) k-vector space filtration $V \supset (\mathrm{rad}\, S)V \supset (\mathrm{rad}\, S)^2 V \supset \ldots$ of V. Thus, the equality $\det_k(T_1) = \det_k(T_2)$ follows from the determinant being multiplicative in exact sequences.

Therefore, to finish the proof of the theorem by proving the first equality of (138), we only need to consider the case where $T \in \mathrm{End}_k(\bar{V}) \subset \mathrm{End}_S(V)$ (this is because $\mathrm{End}_k(\bar{V}) \hookrightarrow \mathrm{End}_S(V) \to \mathrm{End}_k(\bar{V})$ is an isomorphism). In other words, with slightly different language, we can write $T = \bar{T} \otimes \mathrm{id} \in \mathrm{End}_k(\bar{V}) \otimes_k \mathrm{End}_k(S) = \mathrm{End}_k(\bar{V} \otimes_k S) = \mathrm{End}_k(V)$ (though T commutes with S and hence belongs to $\mathrm{End}_S(V)$. In this case, T identifies with a direct sum of $[S : k]$-many copies of $\bar{T}$, so the first equality of (138) is clear.

□

Theorem 27.28. *(i) Suppose A is a finite separable commutative k-algebra, and let $k \hookrightarrow \bar{k}$ be an algebraic closure. Then for all $\alpha \in A$ we have:*

$$\mathrm{tr}_{A/k}(\alpha) = \sum_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A, \bar{k})} \sigma(\alpha), \qquad N_{A/k}(\alpha) = \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A, \bar{k})} \sigma(\alpha).$$

(ii) Suppose K/k is a finite purely inseparable field extension. Then for all $\alpha \in A$ we have:

$$\mathrm{tr}_{K/k}(\alpha) = [K : k]\alpha \quad (= 0 \text{ if } [K : k] > 1), \qquad N_{K/k}(\alpha) = \alpha^{[K:k]}.$$

(iii) For a general field extension K/k,

$$\mathrm{tr}_{K/k}(\alpha) = [K:k]_i \sum_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(K,\bar{k})} \sigma(\alpha), \qquad N_{K/k}(\alpha) = \Big(\prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(K,\bar{k})} \sigma(\alpha) \Big)^{[K:k]_i}.$$

Proof. For (i), base-change to k^s:

$$A \otimes_k k^s \xrightarrow{\cong} \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A,k^s)} k^s,$$

via the map that sends $a \otimes b$ to $(\sigma(a)b)_\sigma$. Thus, multiplication m_a by a on the left-hand side becomes diagonalized on the right-hand side, since in the σ-th coordinate m_a acts as multiplication by $\sigma(a)$. Then just use that the trace and determinant of a diagonal matrix are respectively the sum and the product of its diagonal entries.

Now we come to (ii). Without loss of generality, $\mathrm{char}\, k = p > 0$. We have $a = \alpha^{p^n} \in k$ for some n. We will prove the result for $N_{K/k}$; the result for $\mathrm{tr}_{K/k}$ is analogous. Then $N_{K/k}(\alpha)^{p^n} = N_{K/k}(a) = a^{[K:k]} = (\alpha^{[K:k]})^{p^n}$. Now take p^n-th roots. (For a different approach, base-change to an algebraic closure of k, and use the latter equality of (138)).

(iii) follows from (i) and (ii), since every finite extension can be obtained as a finite separable extension followed by a finite purely inseparable extension (Lemma 27.10), using the transitivity of norm (Theorem 27.27), which we have proved when the base is a field, which in our case is. □

Remark 27.29. The separable case of the above proof tells us that the isomorphism $A \otimes_k K \to \prod_{\sigma \in \mathrm{Hom}_{k\text{-}Alg}(A,K)} K$ for K-split (and hence separable) finite k-algebras A, which we have seen many times by now (e.g., a special case was used in Lemma 27.2), is about diagonalization: this isomorphism lets us diagonalize the multiplication by $a \in A$ on A, but after base-change to K. Note how this ties in with our discussion of Jordan decomposition and maximal separable subalgebras above (Exercise 27.22).

In contrast, when K/k is purely inseparable, $(m_\alpha : x \mapsto \alpha x) \in \mathrm{End}_k(K)$ cannot be diagonalized for $\alpha \in K \backslash k$, because this endomorphism is not semisimple. If one base-changes to an algebraic closure $\bar{k}$, one can see that in $\mathrm{End}_{\bar{k}}(K \otimes_k \bar{k})$, $m_\alpha = m_{\alpha \otimes 1}$ differs from the scalar and hence commuting semisimple endomorphism $m_{1 \otimes \alpha}$ by something nilpotent, hence has $m_{1 \otimes \alpha}$ as the semisimple part of its Jordan decomposition, and hence has the same determinant as $m_{1 \otimes \alpha}$, namely $\alpha^{[K:k]}$ (thus giving a proof of the 'purely inseparable' case of the above theorem). This is essentially the comment made in the above proof about using (138).

28. Lecture 28 – Group cohomology, Artin-Schrier theorem

28.1. **Group homology and cohomology – basic definitions and examples.**

Definition 28.1. Let G be a group, and consider $G\text{-}Mod := \mathbb{Z}[G]\text{-}Mod$, the category of abelian groups with a G-action.

(i) The functor of G-invariants,

$$(-)^G : G\text{-}Mod \rightsquigarrow AbGrp, \quad M \rightsquigarrow M^G := \{m \in M \mid \sigma m = m \,\forall\, \sigma \in G\} \cong \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}, M),$$ [84]

is left exact (e.g., since Hom is left exact), and one defines the group cohomolgy functors $\{H^i(G,-)\}_{i\geqslant 0}$ to be its right derived functors:

$$H^i(G,-) := R^i(M \rightsquigarrow M^G) = \mathrm{Ext}^i_{\mathbb{Z}[G]}(\mathbb{Z}, -).$$

(ii) The functor of G-coinvariants,

$$(-)_G : G\text{-}Mod \rightsquigarrow AbGrp, \quad M \rightsquigarrow M_G := M/\{\sigma m - m \mid \sigma \in G, m \in M\} \cong \mathbb{Z} \otimes_{\mathbb{Z}[G]} M,$$ [85]

is right exact (e.g., since tensor product is right exact), and one defines the group homology functors $\{H_i(G,-)\}_{i\geqslant 0}$ to be its left derived functors:

$$H_i(G,-) := L_i(M \rightsquigarrow M_G) = \mathrm{Tor}_i^{\mathbb{Z}[G]}(\mathbb{Z}, -).$$

Note that, in particular, $H^0(G,-) = (-)^G$ and $H_0(G,-) = (-)_G$.

We will later (in Subsection 28.5) introduce a standard projective resolution to compute the $H^i(G,-)$ and $H_i(G,-)$, but first let us study it as far as possible without such a systematic/standard resolution.

Notation 28.2. We recall notation from our discussion of representation theory: the augmentation ideal in $\mathbb{Z}[G]$ is the kernel of the augmentation map (a ring homomorphism) $\varepsilon : \mathbb{Z}[G] \to \mathbb{Z}$ that sends each $\sigma \in G$ to $1 \in \mathbb{Z}$. Note that, temporarily writing R for $\mathbb{Z}[G]$ and $I \subset \mathbb{Z}[G]$ for the augmentation ideal, we can also describe M^G and M_G (functorially) as follows, and the following description may be used without further mention:

$$M^G = \{m \in M \mid Im = 0\} \cong \mathrm{Hom}_R(R/I, M), \qquad M_G = M/(IM) \cong (R/I) \otimes_R M,$$

where $\mathbb{Z} \cong R/I$ is viewed as a left R-module in the first description, and as a right R-module in the second.

Proposition 28.3. *If the group G acts trivially on the abelian group M, then:*

(i) $H^1(G,M) = \mathrm{Hom}(G,M) = \mathrm{Hom}(G^{ab}, M)$.
(ii) $H_1(G,M) = G^{ab} \otimes_{\mathbb{Z}} M$.

Remark 28.4. The proof will use the following standard facts we have seen before:

[84] where $\mathbb{Z}$ is thought of as a left $\mathbb{Z}[G]$-module with the trivial G-action
[85] where $\mathbb{Z}$ is thought of as a right $\mathbb{Z}[G]$-module with the trivial G-action

(i) Recall from HW 7, Problem 3(iii), that for a left ideal I of a ring R, we have:

$$\mathrm{Tor}_1^R(R/I, M) \cong \ker(I \otimes_R M \overset{\text{mult.}}{\to} IM \subset M).$$

Indeed, this follows from the long exact sequence for Tor associated to the short exact sequence $0 \to I \to R \to R/I \to 0$:

$$0 = \mathrm{Tor}_1^R(R, M) \to \mathrm{Tor}_1^R(R/I, M) \to I \otimes_R M \to R \otimes_R M = M \to (R/I) \otimes_R M \to 0.$$

(ii) We claim

$$\begin{aligned}\mathrm{Ext}^1_R(R/I, M) &= \mathrm{coker}(\mathrm{Hom}_R(R, M) \overset{\text{restriction}}{\to} \mathrm{Hom}_R(I, M)) \\ &\cong \mathrm{Hom}_R(I, M)/\{(i \mapsto im) \mid m \in M\}.\end{aligned}$$

To see this, apply the long exact sequence for Ext to $0 \to I \to R \to R/I \to 0$, to get

$$0 \to \mathrm{Hom}_R(R/I, M) \to \mathrm{Hom}_R(R, M) \to \mathrm{Hom}_R(I, M) \to \mathrm{Ext}^1_R(R/I, M) \to \mathrm{Ext}^1_R(R, M) = 0,$$

and the claim follows easily.

We will need one more preparatory observation for the proof of Proposition 28.3:

Lemma 28.5. *Let $I \subset \mathbb{Z}[G]$ be the augmentation ideal. Then there is an isomorphism of groups $G^{ab} \cong I/I^2$.*

Proof. We have a map $G \to I/I^2$, $g \mapsto (g-1) + I^2$. This is a group homomorphism, because $gh - 1 = (g-1) + (h-1) + (g-1)(h-1) \in (g-1) + (h-1) + I^2$. Since (it is easy to see that) $I = \mathrm{Span}_{\mathbb{Z}}(\{g - 1 \mid g \in G\})$, this shows that $G \to I/I^2$ is surjective as well, and hence factors through a surjective group homomorphism $G^{ab} \to I/I^2$.

In the other direction, note that $\{g - 1 \mid 1 \neq g \in G\}$ is actually a $\mathbb{Z}$-module basis for I. Hence we have an abelian group homomorphism $I \to G^{ab}$, sending each $g-1$, $g \neq 1$, to the image of g in G^{ab}. This homomrphism sends $(g-1)(h-1) = (gh-1) - (g-1) - (h-1)$ to the image of $gh \cdot g^{-1} \cdot h^{-1}$, i.e., to $1 \in G^{ab}$, so this gives a well-defined homomorphism of abelian groups $I/I^2 \to G^{ab}$. It is easy to see that the homomorphisms $G^{ab} \to I/I^2$ and $I/I^2 \to G^{ab}$ that we have defined are two-sided inverses to each other.

Note that G was not assumed to be finite anywhere in the argument. □

Exercise 28.6. In contrast, show that if I is the augmentation ideal in $k[G]$, with G finite and k a field with $(\mathrm{char}\, k, \#G) = 1$, then $I = I^2$.
Hint: Tensor with k and use the right exactness of the tensor product. An alternate approach is to note that I is the product of some $\mathrm{End}_{D_i}(V_i)$ (why?), and hence a ring with the induced multiplication, and the square of the unit ideal is the unit ideal.

Proof of Proposition 28.3. Let $R = \mathbb{Z}[G]$, and let $I \subset R$ denote the augmentation ideal, so that $R/I \cong \mathbb{Z}$. For (i), note that by Remark 28.4, we have:

$$H^1(G, M) = \mathrm{Ext}^1_R(R/I, M) \cong \mathrm{Hom}_R(I, M)/\{(i \mapsto im) \mid m \in M\} \cong \mathrm{Hom}_R(I/I^2, M)/0,$$

where in the last equality we used that I annihilates M, both for the numerator and the denominator.

Hence, using Lemma 28.6 and the fact that $R/I = \mathbb{Z}$ via the augmentation map,

$$H^1(G, M) \cong \mathrm{Hom}_{R/I}(I/I^2, M) \cong \mathrm{Hom}_{\mathbb{Z}}(G^{ab}, M) \cong \mathrm{Hom}_{AbGrp}(G, M),$$

This gives (i), and we move to (ii). Again using that $IM = 0$, we get

$$H_1(G, M) = \mathrm{Tor}_1^R(R/I, M) = \ker(I \otimes_R M \to IM) \cong I \otimes_R M.$$

Since I annihilates M, we get:

$$I \otimes_R M \cong (I \otimes_R (R/I)) \otimes_{R/I} M \cong (I/I^2) \otimes_{\mathbb{Z}} M \cong G^{ab} \otimes_{\mathbb{Z}} M,$$

again using Lemma 28.6. □

Remark 28.7. In the above proof, we approached $H^i(G, M) = \mathrm{Ext}^i_{\mathbb{Z}[G]}(\mathbb{Z}, M)$ by using the long exact sequence associated to Ext in the first argument (we used $0 \to I \to R \to R/I \to 0$), though technically Definition 28.1 involves computing using an injective $\mathbb{Z}[G]$-resolution of M, which gives long exactness in the second argument. Thus, we have already used that the two ways of defining Ext – using a resolution of either argument – are equivalent. Similarly with $H_i(G, M)$.

Example 28.8. Suppose $G = \langle t \mid t^n = 1 \rangle$ is a cyclic group of order n. Let us compute $H^1(G, M)$ for any G-module M. We recall the following computation from the midterm examination. If R is a commutative ring, and $a, b \in R$ are such that $R[b] := \mathrm{Ann}_R(b)$ equals aR and $R[a] := \mathrm{Ann}_R(a)$ equals bR, then note that we have the following resolution of R/a.

$$\cdots \to R \xrightarrow{\times a} R \xrightarrow{\times b} R \xrightarrow{\times a} \ldots \xrightarrow{\times a} R \xrightarrow{\times b} R \xrightarrow{\times a} R \to R/a \to 0.$$

Homming this into M and tensoring this into M respectively, but removing the contribution from R/a (as that is not part of the complex that computes the homology/cohomology), we get the following complexes:

$$0 \to M \xrightarrow{\times a} M \xrightarrow{\times b} M \xrightarrow{\times a} M \xrightarrow{\times b} \ldots, \qquad \text{and} \qquad \ldots \xrightarrow{\times a} M \xrightarrow{\times b} M \xrightarrow{\times a} M \to 0.$$

Taking cohomology/homology, we get
(139)

$$\mathrm{Ext}^i_R(R/a, M) \cong \begin{cases} M[a], & \text{if } i = 0, \\ M[b]/aM, & \text{if } i > 0 \text{ is odd, and} \\ M[a]/bM, & \text{if } i > 0 \text{ is even} \end{cases} \quad \mathrm{Tor}_i^R(R/a, M) \cong \begin{cases} M/aM, & \text{if } i = 0, \\ M[a]/bM, & \text{if } i > 0 \text{ is odd, and} \\ M[b]/aM, & \text{if } i > 0 \text{ is even} \end{cases}$$

We will apply (139) with $R = \mathbb{Z}[G]$ (which is commutative in this case), noting that $a = t - 1$ generates the augmentation ideal I, and with $b = 1 + t + \cdots + t^{n-1}$. It is easy to check the conditions $R[a] = bR$ and $R[b] = aR$. Write $N : M \to M$ for the action of b – note that this looks like a 'norm', from what we saw at the end of Lecture 27.

N clearly factors through $M \to M_G$, and has image in M^G, and hence induces a map $\bar{N} : M_G \to M^G$. Applying (139), and noting that $M^G = M[a]$ and $\ker(\bar{N}) = M[b]/aM$, it follows that for any (possibly nontrivial) G-module M:

$$H^i(G,M) = \begin{cases} M^G, & \text{if } i = 0, \\ \ker(\bar{N}), & \text{if } i > 0 \text{ is odd, and} \\ \operatorname{coker}(\bar{N}), & \text{if } i > 0 \text{ is even} \end{cases} \quad H_i(G,M) = \begin{cases} M_G, & \text{if } i = 0, \\ \operatorname{coker}(\bar{N}), & \text{if } i > 0 \text{ is odd, and} \\ \ker(\bar{N}), & \text{if } i > 0 \text{ is even} \end{cases}.$$

Exercise 28.9. If $G \cong \mathbb{Z} = \langle t \rangle$ is instead an infinite cyclic group, show:

$$H^i(G,M) \cong \begin{cases} M^G, & \text{if } i = 0, \\ M_G, & \text{if } i = 1, \end{cases} \quad H_i(G,M) \cong \begin{cases} M_G, & \text{if } i = 0, \\ M^G, & \text{if } i = 1 \end{cases}.$$

Note: It can help if you note that in this case $\mathbb{Z}[G] = \mathbb{Z}[\mathbb{Z}] \cong \mathbb{Z}[t, t^{-1}]$.

28.2. **Restriction and corestriction.** In this subsection, we would like to prove the following:

Theorem 28.10. *For any finite group G and an abelian group M, multiplication by $\#G$ annihilates $H_i(G,M)$ and $H^i(G,M)$ for all $i > 0$.*

A partial motivation for this theorem is that it gives an alternate proof of Maschke's theorem – though seemingly more complicated, this proof throws light into the failure of Maschke's theorem in bad characteristic (and explains why group cohomology sort of measures this failure):

Alternate proof of Maschke's theorem. The proof of Maschke's theorem boiled down to the assertion that when G is a finite group and k is a field with $(\operatorname{char} k, \#G) = 1$, then $V \rightsquigarrow V^G$ is exact on $Rep_k(G)$. [86] Via the forgetful functor $Rep_k(G) = k[G]\text{-}Mod \rightsquigarrow \mathbb{Z}[G]\text{-}Mod$, and the long exact sequence for $H^*(G,-) = \operatorname{Ext}^*_{\mathbb{Z}[G]}(\mathbb{Z},-)$, this will follow if we show that for each $V \in \operatorname{Ob} Rep_k(G)$, viewing V as a $\mathbb{Z}[G]$-module, we have ($H^1(G,V) = 0$, or) $H^i(G,V) = 0$ for each $i > 0$. But by Theorem 28.10, we know that multiplication by $\#G$ annihilates $H^i(G,V)$. On the other hand, we also know that multiplication by $\#G$ on $H^i(G,V)$ is induced by (multiplication by $\#G$) $:V \to V$ (this is easy, and is the Ext analogue of HW 7 Problem 3(i)), which is an isomorphism as $(\#G, \operatorname{char} k) = 1$. Thus, (multiplication by $\#G$) $:H^i(G,V) \to H^i(G,V)$ is an isomorphism as well as the 0 map, so $H^i(G,V) = 0$. □

To prove Theorem 28.10, it will help to introduce the functors of restriction and corestriction for group homology and cohomology, which are anyway important. For our preparation, we will not assume that G is finite.

[86] This allowed us to get a complement to any subrepresentation $W \subset V$, by getting a section for the surjective map $V \to V/W =: W'$ of representations of G over k: applying the exactness of $V \rightsquigarrow V^G$ to the clearly surjective map $\operatorname{Hom}_k(W',V) \twoheadrightarrow \operatorname{Hom}_k(W',W')$ of representations of G, we obtained that $\operatorname{Hom}_G(W',V) \to \operatorname{Hom}_G(W',W')$ was surjective, and any inverse image of $\mathrm{id} \in \operatorname{Hom}_G(W',W')$ splits $V \to W'$.

Being left and right derived functors, the group homology and cohomology functors are universal δ-functors from $\mathbb{Z}[G]$-Mod to $AbGrp$, $(\{H_i(G,-)\}_i, \{\delta_i\}_i)$ and $(\{H^i(G,-)\}_i, \{\delta^i\}_i)$. Let $H \to G$ be a group homomorphism; viewing G-modules also as H-modules, we have δ-functors $(\{H_i(H,-)\}_i, \{\delta_i\}_i)$ and $(\{H^i(H,-)\}_i, \{\delta^i\}_i)$, again from $\mathbb{Z}[G]$-Mod to $AbGrp$. Thus, on the category $\mathbb{Z}[G]$-Mod of G-modules, we have natural transformations

$$(-)_H \to (-)_G \qquad \text{and} \qquad (-)^G \to (-)^H,$$

consisting of the $M_H \to M_G$ and the $M^G \hookrightarrow M^H$ figuring in the obvious factorizations

$$(M \to M_G) = (M_H \to M_G) \circ (M \to M_H), \quad \text{and} \quad (M^G \hookrightarrow M) = (M^H \hookrightarrow M) \circ (M^G \hookrightarrow M^H).$$

Definition 28.11. Thus, with the above notation, by Grothendieck's theorem on universal δ-functors, we get morphisms of homological/cohomological δ-functors on $\mathbb{Z}[G]$-Mod: corestriction in group homology from H to G,

$$\mathrm{Cores}_H^G : (\{H_i(H,-)\}_i, \{\delta_i\}_i) \to (\{H_i(G,-)\}_i, \{\delta_i\}_i),$$

and restriction in group cohomology from G to H:

$$\mathrm{Res}_H^G : (\{H^i(G,-)\}_i, \{\delta^i\}_i) \to (\{H^i(H,-)\}_i, \{\delta^i\}_i).$$

For us, these notions will only be of interest when $H \to G$ is the inclusion of a subgroup.

In particular we have restriction maps $H^i(G,M) \to H^i(H,M)$ for each i, functorial in M.

The proof of the above theorem uses a corestriction for group cohomology as well, and a restriction for group homology as well. These are defined when $H \subset G$ is a subgroup of finite index. Then we have natural transformations in the opposite direction, i.e., from $(-)_G$ to $(-)_H$ and from $(-)^H$ to $(-)^G$ as well: for all $m \in M$, where M is an abelian group with a G-action,

$$N_{G/H} : M_G \ni (\text{image of } m \in M) \mapsto \sum_{s \in [H\backslash G]} (\text{image of } s \cdot m \text{ under } M \to M_H) \in M_H$$

(don't be like me, check carefully that this is well-defined), and

$$N_{G/H} : M^H \ni m \mapsto \sum_{s \in [G/H]} sm \in M^G,$$

where $[H\backslash G]$ and $[G/H]$ are sets of representatives for $H\backslash G$ and G/H, respectively.

But to get morphisms of δ-functors from here, we need the following:

Lemma 28.12. *When H is a subgroup of finite index in G, $(\{H_i(H,-)\}_i, \{\delta_i\}_i)$ and $(\{H^i(H,-)\}_i, \{\delta^i\}_i)$ are are effaceable/coeffaceable on $\mathbb{Z}[G]$-Mod.*

Proof. For any $\mathbb{Z}[G]$-module M, we have morphisms $P \twoheadrightarrow M \hookrightarrow I$ of $\mathbb{Z}[G]$-modules, with P projective and I injective. It is enough to show that for $i > 0$, we have $H_i(H,P) = 0$ and $H^i(H,M) = 0$. The former is true because P is also projective as a $\mathbb{Z}[H]$-module. For the latter, it is enough to show that I is injective as a $\mathbb{Z}[H]$-module, i.e., $\mathrm{Hom}_{\mathbb{Z}[H]}(-,I)$ is exact. But this is because, by Frobenius reciprocity, $\mathrm{Hom}_{\mathbb{Z}[H]}(-,I)$ is the composite

of $\mathrm{Ind}_H^G = \mathbb{Z}[G] \otimes_{\mathbb{Z}[H]} -$ and $\mathrm{Hom}_{\mathbb{Z}[G]}(-, I)$, both of which are exact (the former is exact because $\mathbb{Z}[G]$ is projective as a $\mathbb{Z}[H]$-module). □

Exercise 28.13. Give an alternate proof of the above lemmas by showing that $H_i(H, -) = \mathrm{Tor}_i^{\mathbb{Z}[G]}(\mathbb{Z}[H\backslash G], -)$, and $H^i(H, -) = \mathrm{Ext}^i_{\mathbb{Z}[G]}(\mathbb{Z}[G/H], -)$.

Notation 28.14. Therefore, Grothendieck's theorem still applies, and with the above notation (in particular $H \subset G$ is of finite index), gives us a morphism of homological δ-functors on $\mathbb{Z}[G]$-Mod:

$$\mathrm{Res}_H^G : (\{H_i(G, -)\}_i, \{\delta_i\}_i) \to (\{H_i(H, -)\}_i, \{\delta_i\}_i),$$

called restriction for group homology from G to H, and a morphism of cohomological δ-functors on $\mathbb{Z}[G]$-Mod:

$$\mathrm{Cores}_H^G : (\{H^i(H, -)\}_i, \{\delta^i\}_i) \to (\{H^i(G, -)\}_i, \{\delta^i\}_i),$$

called the corestriction for group cohomology from H to G. Note that we have only defined these when $H \subset G$ and $[G : H]$ is finite.

Proposition 28.15. *Let $H \subset G$ be a subgroup of finite index. Then for all i and all $M \in \mathrm{Ob}\, G$-Mod,*

$$\mathrm{Cores}_H^G \circ \mathrm{Res}_H^G : H^i(G, M) \to H^i(G, M)$$

is given by multiplication by $\#G$. A similar statement applies to group homology.

Proof. We will prove the assertion involving group cohomology; the assertion involving group homology is similar. For $i = 0$, this is just saying that the map

$$M^G \hookrightarrow M^H \xrightarrow{m \mapsto \sum_{s \in [G/H]} sm} M^G$$

is multiplication by $[G : H]$, which is clear. The general case follows because morphisms from universal δ-functors are completely determined at the $i = 0$ level, and multiplication by $\#(G/H)$ is trivially a morphism of δ-functors.. □

Proof of Theorem 28.10. For the assertion involving group cohomology, apply Proposition 28.15 with H equal to the trivial subgroup of G, and use that in this case $H^i(H, M) = 0$ for all $i > 0$ (e.g., taking fixed points under $\{1\}$ is exact, and hence has vanishing derived functors). The proof of the assertion involving group homology is similar. □

28.3. **Induced modules and Shapiro's lemma.** If $H \subset G$ is a subgroup, we have already introduced the functors $\mathrm{Ind}_H^G : \mathbb{Z}[H]$-$Mod \rightsquigarrow \mathbb{Z}[G]$-$Mod$ and $\mathrm{coInd}_H^G : \mathbb{Z}[H]$-$Mod \rightsquigarrow \mathbb{Z}[G]$-$Mod$ in Lecture 7. Since these are involved in some nice adjointness properties, we can ask if these properties reflect in the world of group homology and cohomology, which is indeed the case:

Theorem 28.16 (Shapiro's lemma)**.** *(i) (Shapiro's lemma for group cohomology) For all $i \in \mathbb{N}$, we have isomorphisms functorial in H-modules M:*

$$H^i(G, \mathrm{coInd}_H^G M) \to H^i(H, M),$$

obtained as a composite $H^i(G, \mathrm{coInd}_H^G M) \to H^i(H, \mathrm{coInd}_H^G M) \to H^i(H, M)$, the latter induced by the morphism $\mathrm{coInd}_H^G M \to M$, and the former given by restriction for group cohomology.

(ii) (Shapiro's lemma for group homology) For all $i \in \mathbb{N}$, we have isomorphisms functorial in H-modules M:

$$H_i(H, M) \to H_i(G, \mathrm{Ind}_H^G M),$$

obtained as the composite $H_i(H, M) \to H_i(H, \mathrm{Ind}_H^G M) \to H_i(G, \mathrm{Ind}_H^G M)$, the former induced by $M \to \mathrm{Ind}_H^G M$, and the latter given by corestriction for group homology.

Remark 28.17. Recall also that coInd_H^G and Ind_H^G are functorially isomorphic when $[G : H]$ is finite.

Partial proof of Theorem 28.16. Only the existence of such isomorphisms will be proved (and that is all that we will need for Artin-Schreier theory below). It will be left to it to the interested readers (if any) to fill in the justification for their descriptions, by making the relevant HW 7 problems more precise.

Let $R = \mathbb{Z}[H]$ and $S = \mathbb{Z}[G]$. Since S is projective as a left R-module, we know from HW 7, problem 5(ii)(b) that for all left R-modules N and left S-modules M, we have for all $i \geqslant 0$:

$$\mathrm{Ext}_R^i(N, M) \cong \mathrm{Ext}_S^i(N, \mathrm{Hom}_R(S, M)). \tag{140}$$

Letting N be $\mathbb{Z}$ as a left S-module, the above gives an isomorphism $H^i(H, M) \to H^i(G, \mathrm{coInd}_H^G M)$. Unfortunately the description of this map was not explicitly given in HW 7 probem 5(ii)(b), but you can check that it agrees with the one given in the proposition.

This gives (i). (ii) is similar, using HW 7, problem 5(ii)(c), which says that whenever S is projective as a right R-module (which is the case in our situation), for all right S-modules N and left R-modules M, we have for all $i \geqslant 0$:

$$\mathrm{Tor}_i^S(N, S \otimes_R M) \xrightarrow{\cong} \mathrm{Tor}_i^R(N, M).$$

□

28.4. **The Artin-Schrier theorem.**

Proposition 28.18 (Hilbert's Theorem 90, additive form)**.** *Let K/k be a finite Galolis extension of fields. Then $H^i(\mathrm{Gal}(K/k), K) = H_i(\mathrm{Gal}(K/k), K) = 0$ for all $i > 0$.*

Proof. Let $G := \mathrm{Gal}(K/k)$. Recall that by the normal basis theorem, K/k is, as a $k[G]$-module, free (and even isomorphic to $k[G]$). Thus, as a G-module, we have identifications $K = k[G] = \mathrm{Maps}(G, k) = \mathrm{coInd}_{\{1\}}^G k = \mathrm{Ind}_{\{1\}}^G k$, where k is given the trivial action of the trivial group $\{1\}$. Therefore, for $i > 0$, we have by Shapiro's lemma (Theorem 28.16),

$$H^i(G, K) \cong H^i(k, \mathbb{1}) = 0, \qquad \text{and} \qquad H_i(G, K) \cong H_i(\{1\}, k) = 0.$$

□

Definition 28.19. A field extension K/k is called cyclic (resp., abelian) if it is Galois and satisifies that $\mathrm{Gal}(K/k)$ is cyclic (resp., abelian).

Corollary 28.20. *Let K/k be a finite cyclic extension, with $\mathrm{Gal}(K/k) = \langle\sigma\rangle$. For $\beta \in K$, we have $\mathrm{tr}_{K/k}(\beta) = 0$ if and only if there exists $\alpha \in K$ such that $\beta = \sigma(\alpha) - \alpha$.*

Remark 28.21. If β is of the form $\sigma(\alpha) - \alpha$, it is immediate that $\mathrm{tr}_{K/k}(\beta) = 0$. It is the converse that is non-obvious.

Proof of Corollary 28.20. Recall the description $H^i(\mathrm{Gal}(K/k), K) = \ker(\bar{N})$ for i odd, from and in the notation of Example 28.8, where we let $G = \mathrm{Gal}(K/k)$ and $M = K$. Applying the additive form of Hilbert's Theorem 90 (Proposition 28.18), we get that $\bar{N}$ is injective. But in this case, $\bar{N}$ identifies with the map

$$M_G = K/\mathrm{Span}_{\mathbb{Z}}\{\sigma'(\alpha) - \alpha \mid \sigma' \in G, \alpha \in K\} \overset{\mathrm{tr}_{K/k}}{\to} K^G = k.$$

The injectivity of $\bar{N}$ means that

$$\{\beta \in K \mid \mathrm{tr}_{K/k}\,\beta = 0\} = \mathrm{Span}_{\mathbb{Z}}(\{\sigma'(\alpha) - \alpha \mid \sigma' \in G, \alpha \in K\}) = \mathrm{Span}_{\mathbb{Z}}(\{\sigma(\alpha) - \alpha \mid \alpha \in K\}),$$

where the last step is an easy consequence of the fact that α is a generator of the cyclic group G. □

Exercise 28.22. Prove Corollary 28.20 by instead considering

$$\alpha = (\mathrm{tr}_{K/k}\,\theta)^{-1}(\beta\sigma(\theta) + (\beta + \sigma(\beta))\sigma^2(\theta) + \cdots + (\beta + \sigma(\beta) + \cdots + \sigma^{n-2}(\beta))\sigma^{n-1}(\theta)),$$

for some $\theta \in K$ such that $\mathrm{tr}_{K/k}(\theta) \neq 0$ (such a θ exists by the separability of K/k, since in that case we have seen in Lecture 24 that the trace form is nondegenerate).

Theorem 28.23. *Let k be a field of characteristic $p > 0$.*

- *(i) Let K/k be a finite cyclic extension. Then there exists $\alpha \in K$ such that $K = k(\alpha)$, and such that α satisfies an equation of the form $x^p - x - a = 0$ for some $a \in k$.*
- *(ii) Conversely, given a polynomial $f(x) = x^p - x - a$ with $a \in k$, one of the following holds:*
  - *(a) f has a root in k, in which case all roots of f are in k; or*
  - *(b) f is irreducible, and adjoining a root α of f to k defines a cyclic extension $k[\alpha]/k$ of degree p.*

Proof. We first prove (i). Let $\mathrm{Gal}(K/k) = \langle\sigma\rangle$. Since $\mathrm{tr}_{K/k}(1) = [K : k](1) = p(1) = 0$, Corollary 28.20 implies that there exists $\alpha \in K$ such that $\sigma(\alpha) - \alpha = 1$.

Note that

$$\sigma(\alpha^p) - \alpha^p = (\alpha + 1)^p - (\alpha + 1) = \sigma(\alpha) - \alpha,$$

so that $a := \alpha^p - \alpha$ is fixed by σ and hence belongs to $K^{\mathrm{Gal}(K/k)} = k$. Therefore α is a root of $x^p - x - a \in k[x]$. Since $\alpha \in K\backslash k$, and $[k[\alpha] : k]$ divides p, we have $[k[\alpha] : k] = p$. This proves (i), though let us also note as an aside that the σ-orbit of α is $\alpha, \alpha+1, \ldots, \alpha+(p-1)$.

Now let us prove (ii), so let $f(x) = x^p - x - a$. If α is a root of f in any extension of k, note that so are $\alpha, \alpha + 1, \ldots, \alpha + p - 1$, and all these are distinct. This already implies that if f has a root in k, then all roots of f are in k. Hence, suppose f does not have a root in k.

Let K/k be the extension obtained by adjoining a root α of f, so that $K = k[\alpha]$, and $[K : k] = d \leqslant p$. It is enough to show that K/k is Galois of degree exactly p, which will also give the irreducibility of f. The argument of the previous paragraph shows that f splits completely in K, so that K/k is normal. It is also separable – since the roots of f are all distinct – and hence Galois. Any nontrivial $\sigma \in \mathrm{Gal}(K/k)$ sends α to $\alpha + i$ for some $0 < i < p$. Since $(p, i) = 1$, for each $0 < j < p$, $\mathrm{Gal}(K/k)$ also contains an automorphism that sends α to $\alpha + j$. This implies that $\#\,\mathrm{Gal}(K/k) \geqslant p = \deg f$, so that $[K : k] = \mathrm{Gal}(K/k) = p$, forcing f to be irreducible. □

28.5. **The bar resolution.** There is an explicit free resolution of the $\mathbb{Z}[G]$-module $\mathbb{Z}$, which helps us to describe the $H^i(G, M)$ more explicitly.

Definition 28.24. For all $n \in \mathbb{N}$, let B_n be the free $\mathbb{Z}$-module (not the free $\mathbb{Z}[G]$-module) on the $(n+1)$-tuples $(\sigma_0, \ldots, \sigma_n) \in G^{n+1}$ (in particular B_0 identifies with $\mathbb{Z}[G]$), made into a $\mathbb{Z}[G]$-module by

$$\sigma \cdot (\sigma_0, \ldots, \sigma_n) = (\sigma\sigma_0, \ldots, \sigma\sigma_n),$$

and define, for $n \geqslant 1$, $\partial = \partial_n : B_n \to B_{n-1}$ by

$$\partial(\sigma_0, \ldots, \sigma_n) = \sum_{i=0}^{n} (-1)^i (\sigma_0, \ldots, \hat{\sigma}_i, \ldots, \sigma_n).$$

For $n = 0$, let

$$\partial = \partial_0 : B_0 = \mathbb{Z}[G] \to \mathbb{Z}$$

be given by the augmentation map.

It is readily checked (exercise) that the B_n and the ∂_n form a complex $(B_\bullet, \partial_\bullet)$. It is also immediate that each B_n is free as a $\mathbb{Z}[G]$-module.

Lemma 28.25. *$B_\bullet \to \mathbb{Z}$ is a resolution.*

Proof. First we check the exactness at B_0. The map $\partial_1 : B_1 = \mathbb{Z}[G \times G] \to \mathbb{Z}[G] = B_0$ sends (σ_0, σ_1) to $\sigma_1 - \sigma_0$. Thus the image of ∂_1 is the augmentation ideal of $\mathbb{Z}[G]$, so that $\mathrm{coker}\,\partial_1$ identifies with $\mathbb{Z}$ via the augmentation map, as desired.

Now let us prove exactness at $n \geqslant 1$. For $n \geqslant 0$, define $H_n : B_n \to B_{n+1}$, by

$$H_n(\sigma_0, \ldots, \sigma_n) = (1, \sigma_0, \ldots, \sigma_n).$$

Check that for $n \geqslant 1$ we have $\partial_{n+1} \circ H_n + H_{n-1} \circ \partial_n = \mathrm{id}_{B_n}$, which immediately gives that for all $v \in \ker \partial_n$, we have

$$v = \partial_{n+1}(H_n(v)) + H_{n-1}(\partial_n(v)) = \partial_{n+1}(H_n(V)) \subset \mathrm{im}(\partial_{n+1}).$$

□

Exercise 28.26. The following may be a better way of writing out the above proof. Throw in a $\mathbb{Z} = B_{-1}$ into the resolution, define an H_{-1}, and interpret the above result as the identity map of the resulting complex being null homotopic.

A resolution involving an explicit $\mathbb{Z}[G]$-basis, on the other hand, would replace B_n by something that has n explicit generators over $\mathbb{Z}[G]$.

Definition 28.27. Let, for $n \geqslant 0$, B'_n be the free (left) $\mathbb{Z}[G]$-module on the set of n-tuples $(\sigma_1, \ldots, \sigma_n)$ with each $\sigma_i \in G$. Here, by convention, B'_0 is understood to be $\mathbb{Z}[G]$, a free left $\mathbb{Z}[G]$-module on the empty tuple. For $n \geqslant 0$, we have an isomorphism $B'_n \to B_n$ of left $\mathbb{Z}[G]$-modules, given on the standard basis of B'_n by

$$(\sigma_1, \ldots, \sigma_n) \mapsto (1, \sigma_1, \sigma_1\sigma_2, \ldots, \sigma_1\sigma_2 \ldots \sigma_n).$$

We transport the maps $\partial_n : B_n \to B_{n-1}$ to maps $\partial'_n : B'_n \to B'_{n-1}$, and the map $B_0 \to \mathbb{Z}$ to a map $B'_0 \to \mathbb{Z}$, via these isomorphisms, and form the resolution $(B'_\bullet, \partial'_\bullet)$ of the left $\mathbb{Z}[G]$-module $\mathbb{Z}$. Note that $B'_0 = \mathbb{Z}[G] \to \mathbb{Z}$ identifies with the augmentation map.

Let us explicitly compute the ∂'_n. We have on B_n:

$$\partial_n(1, \sigma_1, \sigma_1\sigma_2, \ldots, \sigma_1\sigma_2 \ldots \sigma_n) = (\sigma_1, \sigma_1\sigma_2, \ldots, \sigma_1\sigma_2 \ldots \sigma_n)$$
$$+\Big(\sum_{i=1}^{n-1}(-1)^i(1, \sigma_1, \ldots, \sigma_1\widehat{\sigma_2 \ldots}\sigma_i, \ldots, \sigma_1\sigma_2 \ldots \sigma_n)\Big) + (-1)^n(1, \sigma_1, \ldots, \sigma_1\sigma_2 \ldots \sigma_{n-1}).$$

This implies (please ask me if you don't follow this) that

$$\partial'_n(\sigma_1, \ldots, \sigma_n) = \sigma_1(\sigma_2, \ldots, \sigma_n) + \sum_{i=1}^{n-1}(-1)^i(\sigma_1, \ldots, \sigma_{i-1}, \sigma_i\sigma_{i+1}, \ldots, \sigma_n) + (-1)^n(\sigma_1, \ldots, \sigma_{n-1}).$$

Thus, $H^n(G, M)$ identifies with the quotient $Z^n(G, M)/B^n(G, M)$, where inside the abelian group $C^n(G, M)$ of all maps $f : G \times \cdots \times G \to M$ from a product of n copies of G to M, we have

$$Z^n(G, M) = \ker(C^n(G, M) \overset{-\circ\partial'_n}{\to} C^{n+1}(G, M)) = \Big\{f \mid \forall \sigma_1, \ldots, \sigma_{n+1} \in G,$$
$$\sigma_1 f(\sigma_2, \ldots, \sigma_n) + \sum_{i=1}^{n}(-1)^i f(\sigma_1, \ldots, \sigma_i\sigma_{i+1}, \ldots, \sigma_n) + (-1)^{n+1} f(\sigma_1, \ldots, \sigma_n) = 0\Big\},$$

and

$$B^n(G,M) = C^{n-1}(G,M) \circ \partial'_n = \{(\sigma_1,\ldots,\sigma_n) \mapsto \sigma_1 f(\sigma_2,\ldots,\sigma_{n-1}) + \sum_{i=1}^{n-1}(-1)^i f(\sigma_1,\ldots,\sigma_i\sigma_{i+1},\ldots,\sigma_n) + (-1)^n f(\sigma_1,\ldots,\sigma_{n-1}) \mid f \in C^{n-1}(G,M)\}.$$

Remark 28.28. Let us explicitly write out what this means for $n = i = 0, 1, 2$:

(i) For $i = 0$ (interpreting the terms involving $i-1$ appropriately), one gets $B^0(G,M) = 0$, but $Z^0(G,M) = \{m \in M \mid \forall\, \sigma \in G, \sigma(m) - m = 0\} = M^G$, recovering that $H^0(G,M) = Z^0(G,M)/B^0(G,M)$ equals M^G.

(ii) For $i = 1$, one gets

$$Z^1(G,M) = \{(\sigma \mapsto a_\sigma : G \to M) \mid \forall\, \sigma,\tau \in G, \sigma(a_\tau) - a_{\sigma\tau} + a_\sigma = 0\} \subset C^1(G,M),$$

and $B^1(G,M) = \{\sigma \mapsto (\sigma(m) - m) \mid m \in M\} \subset C^1(G,M)$. Thus, the cocycle condition here is "$a_{\sigma\tau} = a_\sigma + \sigma(a_\tau)$", and the coboundaries are the form $\sigma \mapsto \sigma(m) - m$. (Exercise: verify explicitly that the coboundaries are cocycles).

Note that when the G-action on M is trivial, $Z^1(G,M) = \mathrm{Hom}(G,M)$ and $B^1(G,M) = 0$, recovering the description $H^1(G,M) = \mathrm{Hom}(G,M)$ from Proposition 28.3.

(iii) For $i = 2$, one gets

$$Z^2(G,M) = \{(\sigma,\tau) \mapsto a_{\sigma,\tau} : G\times G \to M \mid \forall\, \sigma_1,\sigma_2,\sigma_3 \in G, \sigma_1(a_{\sigma_2,\sigma_3}) - a_{\sigma_1\sigma_2,\sigma_3} + a_{\sigma_1,\sigma_2\sigma_3} - a_{\sigma_1,\sigma_2} = 0\}$$

inside $C^2(G,M)$, and

$$B^2(G,M) = \{(\sigma_1,\sigma_2) \mapsto \sigma_1 a_{\sigma_2} - a_{\sigma_1\sigma_2} + a_{\sigma_1} \mid ((\sigma \mapsto a_\sigma) : G \to M)\} \subset C^2(G,M).$$

Notation 28.29. Elements of $Z^i(G,M)$ are called i-cocycles, and elements of $B^i(G,M)$ are called i-coboundaries. Two i-cocycles in $Z^i(G,M)$ having the same image in $H^i(G,M)$ are said to be cohomologous to each other. Similarly, we define i-cycles and i-boundaries in the homological context.

Exercise 28.30. (i) When the G-action on M is trivial, use the explicit resolution $(B'_\bullet, \partial'_\bullet)$ to recover the identification $H_1(G,M) = G^{ab} \otimes_{\mathbb{Z}} M$ from Proposition 28.3.

(ii) Assume that G is finite cyclic. Use the explicit description of $H^1(G,M)$ in Remark 28.28(i) to recover the formula for $H^1(G,M)$ given in Example 28.8.

Definition 28.31. The description of $H^1(G,-)$ in the abelian case from Remark 28.28(i) can be adapted to give a definition of $H^1(G,M)$ even when M is a nonabelian group, as follows. Namely, in this case:

(i) We first set

$$Z^1(G,M) = \{(\sigma \mapsto a_\sigma) : G \to M \mid a_{\sigma\tau} = a_\sigma \cdot \sigma(a_\tau)\, \forall\, \sigma,\tau \in G\}$$

– it is a set and not a group – and call its elements 1-cocycles of G in M.

(ii) Two elements $(a_\sigma)_\sigma$ and $(a'_\sigma)_\sigma$ are said to be cohomologous, a relation we indicate by $(a_\sigma)_\sigma \sim (a'_\sigma)_\sigma$, if there exists $b \in M$ such that $a'_\sigma = b^{-1}a_\sigma\sigma(b)$ for all $\sigma \in G$. Check that $\sim$ is an equivalence relation, and that if $(a_\sigma)_\sigma \in Z^1(G, M)$ and $b \in G$, we do have $(\sigma \mapsto b^{-1}a_\sigma\sigma(b)) \in Z^1(G, M)$.

(iii) Since $\sim$ is an equivalence relation, we may and do set $H^1(G, M)$ to be the set $Z^1(G, M)/\sim$ of equivalence classes. Its elements are called the (1-)cohomology classes of G in M. Note that $H^1(G, M)$ is not a group in this case, but it is a pointed set, the 'distinguished point' being the equivalence class (i.e., the cohomology class) of $(a_\sigma)_\sigma$, where $a_\sigma = 1 \in M$ for all $\sigma \in G$. Note that we haven't defined $B^1(G, M)$, but rather replaced it with an equivalence relation.

Note that when M is nonabelian, we use multiplicative notation to describe operations in it; in particular we write 1 instead of 0 for its identity element in such situations.

The definition of $H^1(G, M)$ when M is nonabelian will be relevant in Lecture 29, where I hope to discuss the multiplicative Hilbert's Theorem 90, which says that for all $n \geqslant 1$ and all finite Galois extensions L/K, $H^1(\mathrm{Gal}(L/K), GL_n(L))$ is the trivial (i.e., singleton) pointed set.

Exercise 28.32. (Easy and formal) Let G be a group acting on a (not necessarily abelian) group M. Form the semidirect product $M \rtimes G$, and consider the projection map $p : M \rtimes G \to M$.

(i) Show that $Z^1(G, M)$ is in bijection with the set of (group-theoretic) sections $G \to M \rtimes G$ to $p : M \rtimes G \to G$, such that a section $s : G \to M \rtimes G$ and a 1-cocycle $(a_\sigma)_\sigma$ correspond to each other under this bijection if for all $\sigma \in G$, we have $s(\sigma) = (a_\sigma, \sigma)$.

(ii) If sections $s, s' : G \to M \rtimes G$ correspond to cocycles $(a_\sigma)_\sigma, (a'_\sigma)_\sigma$ under the above bijection, show that $(a_\sigma)_\sigma$ and $(a'_\sigma)_\sigma$ are cohomologous to each other if and only if there exists $b \in M$ such that $s' = \mathrm{Int}\, b \circ s$, where $\mathrm{Int}\, b : M \rtimes G \to M \rtimes G$ is given by conjguation by b.

In other words, elements of $H^1(G, M)$ correspond to M-conjugacy classes of sections to $M \rtimes G \to G$.

(iii) When M is abelian, show that the above bijections respect the relevant group structures: e.g., realize the set of sections to $M \rtimes G \to G$ as a group in this case (i.e., when M is abelian), as $Z^1(G, M)$ is, and show that the bijection from the set of sections to $Z^1(G, M)$ is a group homomorphism. Similarly, interpret the bijection from $H^1(G, M)$ to the set of G-conjugacy classes of sections to $M \rtimes G \to G$ as a group isomorphism.

29. Lecture 29 – H^2 and group extensions, Hilbert's theorem 90, basic Kummer theory

Since we are going to write cocycles as $(c_\sigma)_\sigma, (c_{\sigma,\tau})_{\sigma,\tau}$ etc. today, rather than $(a_\sigma)_\sigma, (a_{\sigma,\tau})_{\sigma,\tau}$ etc., let us recall the concrete descriptions of $H^1(G,M)$ and $H^2(G,M)$ from Lecture 28.

(i) For $i=1$, we have

$$Z^1(G,M) = \{(\sigma \mapsto c_\sigma : G \to M) \mid \forall\, \sigma,\tau \in G, \sigma(c_\tau) - c_{\sigma\tau} + c_\sigma = 0\} \subset C^1(G,M),$$

and $B^1(G,M) = \{\sigma \mapsto (\sigma(m) - m) \mid m \in M\} \subset C^1(G,M)$. Thus, the cocycle condition here is "$c_{\sigma\tau} = c_\sigma + \sigma(c_\tau)$", and the coboundaries are the form $\sigma \mapsto \sigma(m) - m$. (Exercise: verify explicitly that the coboundaries are cocycles).

(ii) For $i=2$, one has

$$Z^2(G,M) = \{(\sigma,\tau) \mapsto c_{\sigma,\tau} : G\times G \to M \mid \forall\, \sigma_1,\sigma_2,\sigma_3 \in G, \sigma_1(c_{\sigma_2,\sigma_3}) - c_{\sigma_1\sigma_2,\sigma_3} + c_{\sigma_1,\sigma_2\sigma_3} - c_{\sigma_1,\sigma_2} = 0\}$$

inside $C^2(G,M)$, and

$$B^2(G,M) = \{(\sigma_1,\sigma_2) \mapsto \sigma_1(b_{\sigma_2}) - b_{\sigma_1\sigma_2} + b_{\sigma_1} \mid ((\sigma \mapsto b_\sigma) : G \to M)\} \subset C^2(G,M).$$

29.1. **Some comments on H^1.** Recall that if a group G acts on a not necessarily abelian group M, we have the first group cohomology $H^1(G,M) = Z^1(G,M)/\sim$, where $Z^1(G,M) = \{(c_\sigma)_\sigma : G \to M \mid c_{\sigma\tau} = c_\sigma \cdot \sigma(c_\tau)\}$, and where $(c_\sigma)_\sigma \sim (c'_\sigma)_\sigma$ if and only if there exists $b \in M$ such that $c'_\sigma = b^{-1} c_\sigma \sigma(b)$ for each $\sigma \in G$. This is a pointed set, and if M is an abelian group, this is in bijection the abelian group $Z^1(G,M)/B^1(G,M)$ given earlier (with the distinguished element mapping to 0). Our comments on H^1 will be through the following exercise.

Exercise 29.1. (i) Let the group G act on groups N and M, and let $N \hookrightarrow M$ be an injective homomorphism of groups respecting the action of G. Note that the coset space $X = M/N$ gets an induced action of G. Though $M \to X$ is surjective, $M^G \to X^G$ may not be surjective. However, if $H^1(G,N)$ is trivial (or even if $H^1(G,N) \to H^1(G,M)$ is injective), show that $M^G \to X^G$ is surjective.

Hint: If $x \in X^G$, choose $m \in M$ mapping to x. For all $\sigma \in G$, consider $m^{-1}\sigma(m)$, which is a coboundary for M but only a cocycle for N. If $H^1(G,N)$ is trivial, there exists $n \in N$ such that $n^{-1}\sigma(n) = m^{-1}\sigma(m)$ for all σ, so $mn^{-1} \in M$ is fixed by G and maps to x.

Note: One can make this better an say that there is an 'exact sequence of pointed sets'

$$1 \to N^G = H^0(G,N) \to M^G = H^0(G,M) \to X^G = H^0(G,X) \xrightarrow{\delta} H^1(G,N) \to H^1(G,M),$$

and one can further add in an $H^1(G,X)$ if $N \subset G$ is normal, so that X becomes a group. Here, a sequence $B' \to B \to B''$ of pointed sets is said to be exact if the preimage in B of the distinguished element of B'' is the image of $B' \to B$ (but note that this sort of exact sequence is weaker in its implication than for groups; nevertheless it is still useful).

Note: Here is an application of sorts, though this can be proved more elementarily. If N is a normal subgroup of M, multiplication by $\#G$ is invertible on N, so that $H^1(G, N)$ vanishes by Theorem 28.10 from Lecture 28, and if G acts trivially on N and (M/N), it follows that G acts trivially on M.

(ii) Let $k \hookrightarrow \bar{k}$ be an algebraic closure. One knows that elements of $GL_n(k)$ that are conjugate in $GL_n(k^s)$ are also conjugate in $GL_n(k)$, but this is not true for other groups like the special linear groups SL_n, or symplectic or orthogonal groups. Let M be such a group, where we can talk of $M(k)$ and $M(k^s)$ – for concreteness, we can let $M = SL_n$. The question is when $\gamma, \gamma' \in M(k)$ may be $M(k^s)$-conjugate but not $M(k)$-conjugate. Using the aforementioned exact sequence, show that the $M(k)$-conjugacy classes of elements γ' that are $M(k^s)$-conjugate to γ are in bijection with $\ker(H^1(\mathrm{Gal}(k^s, k), M^\gamma(k^s)) \to H^1(\mathrm{Gal}(k^s/k), M(k^s))$, where $M^\gamma(k^s)$ is the centralizer of γ in $M(k^s)$, and 'kernel' means 'inverse image of the distinguished element'.

Thus, if we know some Galois cohomology vanishing, we may be able to say that certain elements of an algebraic group M that are 'conjugate over k^s' are actually 'conjugate over k', or at least quantify the failure in this being the case.

29.2. H^2 **and group extensions.** One motivation for considering H^2 is that it allows us to construct a generalization of the semidirect product. Let A be an abelian group, and G a group acting on A, so that $A \rtimes G$ is well-defined, with underlying set $A \times G$. How can we get other group structures on the same underlying set $A \times G$?

Notation 29.2. Until Proposition 29.6 below, we fix a group G acting on an abelian group A. Given a function $c : G \times G \to A$ denoted as $(c_{\sigma,\tau})_{\sigma,\tau}$, we let $E = E_c$ be equal to $A \times G$ as a set, but given the binary operation:

$$(a_1, \sigma_1) \cdot (a_2, \sigma_2) = (a_1 + \sigma_1(a_2) + c_{\sigma_1,\sigma_2}, \sigma_1\sigma_2). \tag{141}$$

Example 29.3. When $c : G \times G \to A$ is identically 0, $E = E_c$ is clearly a group: it is the semidirect product $A \rtimes G$, where the semidirect product is defined using the action already fixed.

Of course, the question here is: what are the $c = (c_{\sigma,\tau})_{\sigma,\tau}$ for which $E = E_c$ is a group (i.e., with the multiplication defined as in (141)) above?

Exercise 29.4. Assume that $E = E_c$ is a group. Show that we have an exact sequence of groups

$$1 \to A \to E \to G \to 1. \tag{142}$$

Note: Though *Grp* is not an abelian category, a sequence $1 \to A \to E \to G \to 1$ of homomorphisms of (not necessarily abelian) groups is defined to be exact if $A \to E$ is injective, $E \to G$ is surjective, and the image of $A \to E$ is the kernel of $E \to G$.
Hint: It is clear that $E \to G$ is a group homomorphism, and the kernel is $A \times \{1\}$ with the binary operation $(a_1, 1)(a_2, 1) = (a_1 + a_2 + c_{1,1}, 1)$. Then $(a, 1) \mapsto a + c_{1,1}$ defines a group isomorphism from this group to A.

What are the conditions on $c : (\sigma_1, \sigma_2) \mapsto c_{\sigma_1,\sigma_2}$ that turn E_c in to a group?

Invertibility of actions: It is immediately checked that regardless of what $(\sigma_1, \sigma_2) \mapsto c_{\sigma_1,\sigma_2}$ is, the map $(a, \sigma) \cdot - : A \times G \to A \times G$ is always bijective for all $(a, \sigma) \in E_c$: informally, it is bijective on the 'G'-coordinate, and the extra c_{σ_1,σ_2} on the 'A'-coordinate can be adjusted by modifying the a_2. An analogous assertion applies to 'multiplying from the right'.

Associativity: The associativity condition translates to requiring, for all $(a_1, \sigma_1), (a_2, \sigma_2), (a_3, \sigma_3) \in E_c$, that:

$$a_1 + \sigma_1(a_2) + c_{\sigma_1,\sigma_2} + \sigma_1\sigma_2(a_3) + c_{\sigma_1\sigma_2,\sigma_3} = a_1 + \sigma_1(a_2) + \sigma_1\sigma_2(a_3) + \sigma_1(c_{\sigma_2,\sigma_3}) + c_{\sigma_1,\sigma_2\sigma_3}.$$

It follows that the multiplication is associative if and only if $c = (c_{\sigma,\tau})_{\sigma,\tau}$ is a 2-cocycle.

Identity element: An identity element, if any, is of the form $(a, 1)$; and it is immediately verified that $(a, 1)$ is a left identity (resp., right identity) if and only if $\sigma = 1$ and, for all (a', σ') we have $c_{1,\sigma'} = -a$ (resp., $c_{\sigma',1} = -\sigma'(a)$).

So for a general c, such an a does not exist (e.g., why should $c_{1,\sigma'}$ be independent of σ'?). However:

Exercise 29.5. (i) If $c = (c_{\sigma,\tau})_{\sigma,\tau}$ is a 2-cocycle of G in A, show that there exists a unique $a \in A$ such that for all $\sigma \in G$ we have $c_{1,\sigma} = -a$ and $c_{\sigma,1} = -\sigma(a)$. In other words, show that for all $\sigma \in G$ we have $c_{1,\sigma} = c_{1,1}$, and that $c_{\sigma,1} = \sigma(c_{1,1})$.

(ii) Show that (any general) 2-cocycle $c = (c_{\sigma,\tau})_{\sigma,\tau}$ of G in A is cohomologous to a normalized 2-cocycle $c' = (c'_{\sigma,\tau})_{\sigma,\tau}$, where 'normalized' means having the property that $c'_{\sigma,1} = c'_{1,\sigma} = 0$ for all $\sigma \in G$.
Hint: Let c and c' differ by the coboundary db of any $b = (b_\sigma)_\sigma \in C^1(G, A)$ such that $b_1 = c_{1,1}$.
Note: One advantage of a normalized cocycle is that, if $c \in Z^2(G, A)$ is normalized, checking the identity axiom for E_c as in (141) becomes trivial: $(0, 1) \in A \times G = E$ is an identity under these assumptions. Thus, Exercise 29.4 is even more trivial when c is normalized. Proposition 29.6(ii) below, together with the present exercise, tells us that to study E_c, we can always reduce to the case where c is a normalized 2-cocycle.

Proposition 29.6. *Let a group G act on an abelian group A. Fix a map $c = (c_{\sigma,\tau})_{\sigma,\tau} : G \times G \to A$, and form $E_c = A \times G$ as in Notation 29.2, with the binary operation* (141).

(i) E_c, given the binary operation (141)*, is a group if and only if c is a 2-cocycle.*
(ii) Suppose 2-cocycles $c = (c_{\sigma,\tau})_{\sigma,\tau}, c' = (c'_{\sigma,\tau})_{\sigma,\tau} \in Z^1(G, A)$ are 2-cocycles with the same image in $H^2(G, A)$; say,

$$c(\sigma_1, \sigma_2) - c'(\sigma_1, \sigma_2) = \sigma_1(b_{\sigma_2}) - b_{\sigma_1\sigma_2} + b_{\sigma_1}, \qquad \forall\, \sigma_1, \sigma_2 \in G, \tag{143}$$

where $(b_\sigma)_\sigma \in C^1(G,A)$ is a 1-cochain. Then defining E_c and $E_{c'}$ as above, we have a commutative diagram of homomorphisms of groups

$$\begin{array}{ccccccccc} 1 & \longrightarrow & A & \longrightarrow & E_c & \longrightarrow & G & \longrightarrow & 1\ . \\ & & \| & & \downarrow{\scriptstyle (a,\sigma)\mapsto(a+b_\sigma,\sigma)} & & \| & & \\ 1 & \longrightarrow & A & \longrightarrow & E_{c'} & \longrightarrow & G & \longrightarrow & 1 \end{array} \tag{144}$$

(iii) Assume that c is a 2-cocycle, so that E_c is a group by (i). Then E_c fits into an exact sequence

$$1 \to A \to E \to G \to 1,$$

with the following further property. Since A is abelian, the conjugation action of E on A quotients to an action of G on A; the property is that this action is the action of G on A that we started with.

Proof. (i) follows from the above discussion: to summarize, we have shown that the associativity of the binary operation on E_c is equivalent to c being a 2-cocycle, that c being a 2-cocycle automatically implies the existence of an identity element (see Exercise 29.5(i)), and that left or right 'multiplication' by any (a,σ) is automatically bijective.

For (ii), what one needs to verify is the following, which follows from (144).

$$a_1 + \sigma_1(a_2) + c_{\sigma_1,\sigma_2} + b_{\sigma_1\sigma_2} = a_1 + b_{\sigma_1} + \sigma_1(a_2 + b_{\sigma_2}) + c'_{\sigma_1,\sigma_2}, \qquad \forall\, \sigma_1,\sigma_2 \in G.$$

The existence of the exact sequence as in (iii) has been covered in Exercise 29.4. To show the assertion about the action of G on A, by (ii) and Exercise 29.5(ii), we may and do assume that c is normalized, i.e., that $c_{\sigma,1} = c_{1,\sigma} = 0$ for all σ. Now the claim about the action of G on A follows from the computation

$$(0,\sigma)(a,1) = (\sigma(a),\sigma) = (\sigma(a),1)(0,\sigma),$$

where the first equality used that $c_{\sigma,1} = 0$, and the second that $c_{1,\sigma} = 0$. □

Notation 29.7. (i) Let G be a group and A an abelian group. If no action of G on A is specified, then an extension of G by A refers to a group E together with an exact sequence of groups $1 \to A \to E \to G \to 1$. Any such extension/exact sequence determines an action of G on A, the one obtained by quotienting the conjugation action of E on A.

If an action of G on A is specified, then in that context an extension of G by A will again refer to an exact sequence $1 \to A \to E \to G \to 1$ of groups, but with the additional property that the action of G on A determined by this exact sequence equals the given action of G on A.

(ii) Two extensions $1 \to A \to E \to G \to 1$ and $1 \to A \to E' \to G \to 1$ of G by A are isomorphic if these are the rows of a commutative diagram as in (144), whose middle vertical arrow is some isomorphism of groups $E \to E'$.

(iii) An extension E of G by A is said to be a central extension if the associated action of G on A is trivial. This is easily seen to be equivalent to $A \subset E$ being a central subgroup of E.

Theorem 29.8. *Let a group G act on an abelian group A. Associating to a 2-cocycle c of G in A the set $E_c = A \times G$ together with the binary operation* (141)*, induces a bijection*

$$H^2(G,A) \to \{^{\text{Isomorphism classes of extensions of } G \text{ by } A,}_{\text{for the given action of } G \text{ on } A.} \tag{145}$$

Under this bijection, the 0*-element of $H^2(G,A)$ corresponds to $A \rtimes G$ (so the other elements of $H^2(G,A)$ correspond to extensions $1 \to A \to E \to G \to 1$ that are not semidirect products, or equivalently, such that $E \to G$ does not have a section).*

Proof. The existence of a well-defined map as given follows from Proposition 29.6. The assertion about the image of $0 \in H^2(G,A)$ follows from Example 29.3. It is therefore enough to show that (145) is a bijection.

For this, one constructs a candidate two-sided inverse as follows. Given an extension of E by A inducing the given action of G on A, we choose a set-theoretic section $(\sigma \mapsto s_\sigma) : G \to E$ to $E \to G$, and set $c_{\sigma_1,\sigma_2} = s_{\sigma_1} s_{\sigma_2} s_{\sigma_1\sigma_2}^{-1}$.

Any other set-theoretic section is of the form $bs : G \to E$, where $b : G \to A$, replacing s by which replaces c by a cohomologous 2-cocycle, as one sees by verifying the following computation:

$$b_{\sigma_1} s_{\sigma_1} \cdot b_{\sigma_2} s_{\sigma_2} \cdot (b_{\sigma_1\sigma_2} s_{\sigma_1\sigma_2})^{-1} = b_{\sigma_1} \sigma_1(b_{\sigma_2}) b_{\sigma_1\sigma_2}^{-1} \cdot s_{\sigma_1} s_{\sigma_2} s_{\sigma_1\sigma_2}^{-1}.$$

This gives a candidate map in the other direction; denote it by $E \mapsto c_E$ (adopting some informal notation: c_E is really in $H^2(G,A)$ but we might confuse it with a representative in $Z^2(G,A)$). To finish the proof, we need to show that c_{E_c} is cohomologous to c, and that E_{c_E} is isomorphic to E.

Let us first prove the latter. Note that while constructing this map using the set-theoretic section s, resulting in the 2-cocycle $c = c_E$, $A \times G$ identifies set-theoretically with E via $(a,\sigma) \mapsto as(\sigma)$. Under the inverse of this identification, the multiplication on E is transported to the binary operation on $A \times G$ given by (141), as follows from the computation:

$$a_1 s(\sigma_1) a_2 s(\sigma_2) = a_1 \sigma_1(a_2) s(\sigma_1) s(\sigma_2) = a_1 \sigma_1(a_2) \cdot (s(\sigma_1) s(\sigma_2) s(\sigma_1\sigma_2)^{-1}) \cdot s(\sigma_1\sigma_2).$$

But this is giving exactly the equality $E \cong E_{c_E}$. To finish the proof, it is enough to show that c_{E_c} is cohomologous to E. Start with a 2-cocycle c, which we assume without loss of generality to be normalized, so that $c_{\sigma,1} = c_{1,\sigma} = 0$ for all $\sigma \in G$. Then the computation

$$(0,\sigma_1)(0,\sigma_2) = (c_{\sigma_1,\sigma_2}, \sigma_1\sigma_2) = (c_{\sigma_1,\sigma_2}, 1) \cdot (0, \sigma_1\sigma_2),$$

where the latter equality uses the normalization of c, shows that c_{E_c} is cohomologous to c. □

Exercise 29.9. Show that any two normalized 2-cocycles of G in A (again, $(c_{\sigma,\tau})_{\sigma,\tau}$ is 'normalized' if $c_{\sigma,1} = c_{1,\sigma} = 0$ for all $\sigma \in G$) differ from each other by a normalized 2-coboundary, i.e., by a db, where $b(1) = 0$. Thus, $H^2(G, A)$ can also be described as the quotient of the group of normalized 2-cocycles by that of normalized 2-coboundaries.

A better, more systematic and general, treatment of group cohomology using normalized cocycles is found in Professor Nair's notes, where a normalized bar resolution is defined, using which one can also define the $H^i(G, A)$, for all i, in terms of what are called normalized i-cocycles and normalized i-coboundaries.

29.3. **The Schur-Zassenhaus theorem.**

Theorem 29.10 (Schur-Zassenhaus)**.** *Let G, H be finite groups with $(\#G, \#H) = 1$. Then any exact sequence $1 \to H \to E \to G \to 1$ of groups is split, i.e., we have an isomorphism $E \cong H \rtimes G$.*

Proof. The base case of the induction will be when H is abelian. In this case, by Theorem 29.8, extensions for any given action of G on H are classified by $H^2(G, H)$. Since $(\#G, \#H) = 1$, on one hand, multiplication by $\#G$ on (the abelian group) H is an isomorphism and hence induces one on $H^2(G, H)$. On the other hand, multiplication by $\#G$ induces the 0 map on $H^2(G, H)$ by Theorem 28.10 from Lecture 28. Thus, multiplication by $\#G$ is both an isomorphism and the 0 map on $H^2(G, H)$, so we have $H^2(G, H) = 0$.

Now we will reduce the general case to the abelian case, by induction on $\#H$. Assume the result to be true whenever $\#H$ is smaller. Let $P \subset H$ be a p-Sylow subgroup for some prime p; it is a p-Sylow subgroup of E as well, since $(\#H, \#G) = 1$. First let us reduce to the case where E normalizes P. Since H is normal in E, any p-Sylow subgroup of E, being conjugate to P, is contained in H, from which it follows that the map $E \to G$ remains surjective when restricted to the normalizer $N_E(P) \subset E$ of P in E (since any E-conjugate of P can be H-conjugated to P). Now we have an exact sequence $1 \to H \cap N_E(P) \to N_E(P) \to G \to 1$, and it is enough to find a section to $N_E(P) \to G$ (since $N_E(P) \to G$ is a restriction of $E \to G$). Thus, we have reduced to the case where $E = N_E(P)$, i.e., to the case where P is normalized by E, which we assume now.

Let $Z = Z(P)$ be the center of P. Since H and E normalize P, they normalize Z as well, and we get an exact sequence

$$1 \to H/Z \to E/Z \to G \to 1.$$

By the induction hypothesis (which applies since Z is nontrivial), $E/Z \to G$ splits, so E/Z has a subgroup G' isomorphic to G. If $E' \subset E$ is the inverse image of G' under $E \to E/Z$, then we have an exact sequence

$$1 \to Z \to E' \to G' \cong G \to 1.$$

Since Z is abelian, this exact sequence splits, as desired. □

29.4. **Application to projective representations.** Here is one way 2-cocycles help. Sometimes, we don't have a representation $G \to GL_k(V)$, but only a homomorphism $G \to PGL_k(V)$: or equivalently, a noncanonical map $g \mapsto A_g \in GL_k(V)$ such that $G \to GL_k(V) \to PGL_k(V)$ is a homomorphism. Since a lot of representation theory has been studied and found to be useful, we would like to ask: can the homomorphism $G \to PGL_k(V)$ can be studied using a representation? This would be easy if $G \to PGL_k(V)$ lifts to a homomorphism $G \to GL_k(V)$, but such a lift may not exist. We might like to find the "next best approximation".

Here is one way this sort of a situation can arise. Suppose that (ρ, V) is an irreducible representation of H over an algebraically closed field k, and that G acts on H in such a way that for each $g \in G$, ${}^g\rho := \rho \circ \operatorname{Int} g^{-1} : H \to GL_k(V)$ is isomorphic to ρ. This means that for each $g \in G$, we have a linear map $A_g : V \to V$ such that ${}^g\rho(h) \circ A_g = A_g \circ \rho(h)$ for all $h \in H$ – i.e., A_g intertwines ρ with ${}^g\rho$. It is then easy to see that $A_g A_h$ and A_{gh} both intertwine ρ with ${}^{gh}\rho$. In other words, $(A_{gh})^{-1} \circ A_g A_h$ intertwines V with V, and is hence, by Schur's lemma, a scalar in k. But this is the same as saying that $g \mapsto \bar{A}_g$ is a homomorphism $G \to PGL_k(V)$, where $\bar{A}_g$ is the image of A_g under $GL_k(V) \to PGL_k(V)$.

In general, we may not be able to 'correct' the A_g by scalars so that $g \mapsto A_g$ becomes a homomorphism $G \to GL_k(V)$. In such a situation, how can we still get a representation out of $G \to PGL_k(V)$? One easily verifies that $(g, h) \mapsto A_g A_h A_{gh}^{-1}$ is a 2-cocycle valued in $k^\times \cong Z(GL_k(V))$, and thus gives rise to an element of $H^2(G, k^\times)$, where let G act trivially on $k^\times$. If $A = k^\times$, or if $A \subset k^\times$ is a subgroup containing the image of this cocycle, then we get, by Theorem 29.8, a group extension $1 \to A \to \tilde{G} \to G \to 1$ correpsonding to the class of this cocycle in $H^2(G, A)$, and it is easy to see that there is a unique homomorphism $\tilde{G} \to GL_k(V)$, call it $\tilde{g} \mapsto \tilde{A}_{\tilde{g}}$, such that the following diagram commutes:

$$\begin{array}{ccc} \tilde{G} & \xrightarrow{\tilde{g} \mapsto \tilde{A}_{\tilde{g}}} & GL_k(V) \\ \downarrow & & \downarrow \\ G & \xrightarrow{g \mapsto A_g} & PGL_k(V) \end{array} .$$

Thus, the projective representation $G \to PGL_k(V)$ may not lift to a representation $G \to GL_k(V)$, but it will lift to a representation $\tilde{G} \to GL_k(V)$, where $\tilde{G}$ is a central extension of G by some abelian group A.

A particularly important example is when we have a symplectic space $(W, \langle \cdot, \cdot \rangle)$ over, say $\mathbb{R}$. Then one can form the group H whose underlying set is $\mathbb{R} \times W$, and whose multiplication is given by:

$$(a_1, w_1) \cdot (a_2, w_2) = \Big(a_1 + a_2 + \frac{1}{2}\langle w_1, w_2 \rangle, w_1 + w_2\Big).$$

Then H is called the Heisenberg group associated to W and $\langle \cdot, \cdot \rangle$. By the famous Stone-von Neumann theorem, given a nontrivial (continuous) character $\psi : \mathbb{R} \to \mathbb{C}^\times$, there exists a unique irreducible unitary representation $(\rho = \rho_\psi, V = V_\psi)$ of H on a Hilbert space $V = V_\psi$, with the property that $Z(H) = \mathbb{R} \subset H$ acts on $V = V_\psi$ through ψ.

Note that the symplectic group $G = Sp(W, \langle \cdot, \cdot \rangle)$ acts by automorphisms on H: $g \in G$ sends (a, w) to $(a, g \cdot w)$, which is a group homomorphism since $\langle w_1, w_2 \rangle = \langle gw_1, gw_2 \rangle$. Since G acts as the identity on $Z(H)$, the uniqueness of $(\rho = \rho_\psi, V = V_\psi)$ up to isomorphism implies that ${}^g\rho := \rho \circ \operatorname{Int} g^{-1}$ is isomorphic to ρ for each $g \in G$. One is now in the situation described earlier, and we get $A_g \in GL_{\mathbb{C}}(V)$ as above. One can show that in this case, we don't get a representation of G or $H \rtimes G$ on V – the A_g cannot be chosen so that $A_{g_1 g_2} = A_{g_1} A_{g_2}$ for each g and h – but only one of $H \rtimes \tilde{G}$, where $\tilde{G} \to G$ is a certain central extension of G. In fact, one can show that this $\tilde{G} \to G$ can be chosen to have kernel $\{\pm 1\}$. The resulting $\tilde{G}$ is what is called the metaplectic group, which has a lot of importance in the theory of modular forms.

29.5. **(Multiplicative) Hilbert's Theorem 90.**

Theorem 29.11 (Hilbert's Theorem 90)**.** *Let K/k be a finite Galois extension.*

(i) We have $H^1(\mathrm{Gal}(K/k), K^\times) = 0$ (where $\mathrm{Gal}(K/k)$ has its obvious action on $K^\times$).
(ii) More generally, for any positive integer n, the pointed set $H^1(\mathrm{Gal}(K/k), GL_n(K))$ is trivial (i.e., singleton).

Proof. Though (ii) is a generalization of (i), we will first give a stand-alone proof of (i), which is more concrete though less obviously conceptual.

So let a 1-cocycle $\sigma \mapsto c_\sigma$ of $\mathrm{Gal}(K/k)$ in $K^\times$ be given; it is enough to show that it is a coboundary. By the linear independence of characters, there exists $y \in K$ such that $b := \sum_{\tau \in \mathrm{Gal}(K/k)} c_\tau \tau(y) \neq 0$.

Then for each $\sigma \in \mathrm{Gal}(K/k)$ we have

$$\sigma(b) = \sigma(\sum_\tau c_\tau \tau(y)) = \sum_\sigma \sigma(c_\tau)\sigma\tau(y) = \sum_\sigma c_\sigma^{-1} c_{\sigma\tau} \cdot \sigma\tau(y) = c_\sigma^{-1} b.$$

Therefore, we have $c_\sigma = \sigma(b)^{-1} b$ for each $\sigma \in \mathrm{Gal}(K/k)$, which is to say, $(c_\sigma)_\sigma$ is a coboundary, proving (i).

Now we come to the more general statement, (ii). Again, we start with a cocycle $(c_\sigma)_\sigma$, with each $c_\sigma \in GL_n(K)$. As before, it is enough to show that this cocycle is a coboundary, i.e., that there exists $A \in GL_n(K)$ such that $c_\sigma = \sigma(A)^{-1} A$ for all $\sigma \in \mathrm{Gal}(K/k)$.

Consider $V := K^n$ as a vector space over K with a new action of $\mathrm{Gal}(K/k)$, where the new action of $\sigma \in \mathrm{Gal}(K/k)$ on $v \in V$ will be denoted by $\sigma_{new} \cdot v$:

$$\sigma_{new} \cdot v := c_\sigma(\sigma(v)) \in GL_n(K)(K^n) = K^n.$$

Of course, one needs to check that this is indeed an action: one checks that $\sigma_{new} \circ \tau_{new} = (\sigma\tau)_{new}$ using the cocycle condition (i.e., $c_{\sigma\tau} = c_\sigma \sigma(c_\tau)$), and that $1_{new} = 1$ (since $c_1 = c_1 \cdot 1(c_1) = c_1^2 \in GL_n(K)$, we have $c_1 = 1 \in GL_n(K)$). [87]

[87] Here is another way of checking that this is an action. $GL_n(K) \rtimes \mathrm{Gal}(K/k)$ acts on K^n (I will leave it to you to make sense of the details for this), and recall from an Exercise in Lecture 28 that $\sigma \mapsto (c_\sigma, \sigma)$ is a homomorphism $\mathrm{Gal}(K/k) \to GL_n(K) \rtimes \mathrm{Gal}(K/k)$.

It is immediate that this new action is $\mathrm{Gal}(K/k)$-semilinear, i.e., $\sigma_{new}(av) = \sigma(a)\sigma_{new}(v)$ for all $a \in K$ and $v \in V$. Thus, by Galois descent (Theorem 26.4 from Lecture 26), K^n, with this new $\mathrm{Gal}(K/k)$-action, is isomorphic to $k^n \otimes_k K = K^n$ with the usual Galois action, i.e., there exists a basis $v_1, \ldots, v_n \in K^n = V$ that is fixed by the new action. Let $e_1, \ldots, e_n$ be the standard basis of K^n (fixed by the usual action).

Let $A \in GL_n(K)$ be such that $A^{-1}e_i = v_i$ for $1 \leqslant i \leqslant n$. Then for each $1 \leqslant i \leqslant n$, since $\sigma_{new}(v_i) = v_i$, for each $\sigma \in \mathrm{Gal}(K/k)$ we have $c_\sigma(\sigma(A^{-1}e_i)) = A^{-1}e_i$, so $c_\sigma \circ \sigma(A^{-1}) = A^{-1}$, i.e., $c_\sigma = A^{-1}\sigma(A)$ for each σ, which shows that $(c_\sigma)_\sigma$ is a coboundary, finishing the proof of (ii). □

Remark 29.12. (i) The moral of the above proof of Theorem 29.11 is that Hilbert's Theorem 90 (in its multiplicative form) is basically Galois descent (and thus, in a sense equivalent to Galois theory itself; see the notes for Lecture 26).

(ii) How are the arguments given for (i) and (ii) of Theorem 29.11 above related? In Galois descent, one gets invariant vectors by summing over (rather than averaging over) $\mathrm{Gal}(K/k)$-orbits – see the proof of Proposition 26.6 in the notes for Lecture 26, which was used to prove Galois descent for vector spaces, Theorem 26.4 from Lecture 26. This is exactly what was done in (i) – the $\sum_\tau c_\tau \tau(y)$ is, in terms of notation from the proof of (ii) of the theorem, simply the sum over the $\mathrm{Gal}(K/k)$-orbit of y for the 'new' action of $\mathrm{Gal}(K/k)$. The fact that the choice of y involved linear independence of characters exactly mirrors the use of linear independence of characters in the proof of Proposition 26.6 from Lecture 26.

(iii) I have not thought about whether the additive form of Hilbert's Theorem 90 has a similar interpretation. Perhaps something involving "affine spaces over the K-vector space K", but I am not sure. If you figure this out, please let me know.

(iv) A very vague remark follows, although one that is a very important point about what H^1 does. Let X be a 'mathematical object' over a field k. What are the mathematical objects Y over k with the property that X and Y become isomorphic upon base-change to an extension K/k? Say we have $\varphi : X_K \to Y_K$. Since X and Y may not be isomorphic over k, it can happen that for each $\sigma \in \mathrm{Gal}(K/k)$, ${}^\sigma\varphi := \sigma \circ \varphi \circ \sigma^{-1}$ is different from φ. Consider $\varphi^{-1} \circ {}^\sigma\varphi : X_K \to X_K$, or rather $(c_\sigma := \varphi^{-1} \circ {}^\sigma\varphi \in \mathrm{Aut}(X_K))_{\sigma \in \mathrm{Gal}(K/k)}$. It is easy to see that $(c_\sigma)_\sigma$ is a 1-cocycle of $\mathrm{Gal}(K/k)$ in $\mathrm{Aut}(X_K)$ (after all, "$\varphi^{-1} \circ {}^\sigma\varphi$" looks like a coboundary, except that φ does not belong to $\mathrm{Aut}(X_K)$; nevertheless, this is enough for getting a cocycle). One then shows that assigning the isomorphism class of Y over k to $(c_\sigma)_\sigma$ induces a bijection between $H^1(\mathrm{Gal}(K/k), \mathrm{Aut}(X_K))$, and the '$K/k$-forms of X', i.e., the isomorphism classes of mathematical objects Y over k such that X and Y become isomorphic over K. This seems to be at least reminiscent of, if not the very basis of, what happened with Hilbert's Theorem 90 above: $GL_n(K) = \mathrm{Aut}_K(K^n) = \mathrm{Aut}_k(k^n \otimes_k K)$, so $H^1(\mathrm{Gal}(K/k), GL_n(K))$ should classify something like vector spaces over k that become isomorphic to K^n over K; but Galois descent implies that there can only be one such object over k, namely the k-vector space k^n, so

it follows that $H^1(\mathrm{Gal}(K/k), GL_n(K))$ is trivial (but this is informal; I haven't thought about this well enough to make this vague heuristic precise).

For a more concrete example, Exercise 29.1(ii) exemplifies this philosophy of H^1.

Remark 29.13. What about infinite Galois extensions K/k? In this case, $\mathrm{Gal}(K/k)$ is to be considered with its Krull topology, so it turns out that group cohomology as defined in Lecture 28 (and dealt with above) is not the correct object to consider while dealing with it. However, there is a slight variant of the theory of group cohomology that does satisfactorily apply to this situation, as we now outline.

(i) Rather than look at the category of abelian groups on which $\mathrm{Gal}(K/k)$ acts, one should look at the category of abelian groups viewed as topological groups with the discrete topology, on which $\mathrm{Gal}(K/k)$ acts *continuously* (thus, actions on abelian groups A such that the stabilizer of each element of A in $\mathrm{Gal}(K/k)$ is an open subgroup).

(ii) On this category, we can define the functor of $\mathrm{Gal}(K/k)$-invariants, and derive this functor to get functors that one denotes by $H^i(\mathrm{Gal}(K/k), -)$ (which are thus different from what is denoted by $H^i(\mathrm{Gal}(K/k), -)$ as per the definition in Lecture 28).

(iii) One can show that this new $H^i(\mathrm{Gal}(K/k), A)$ can also be described as $Z^i(G, A)/B^i(G, A)$, but where this time $Z^i(G, A)$ and $B^i(G, A)$ are required to be *continuous* maps $G \times \cdots \times G \to A$ that satisfy the same cocycle or coboundary condition as considered in Lecture 28.

(iv) One can also show that $H^i(\mathrm{Gal}(K/k), A)$ is in a suitable sense the directed colimit of the $H^i(\mathrm{Gal}(K/k)/H, A^H)$ as H varies over the open normal subgroups of A: the groups $\mathrm{Gal}(K/k)/H$ are finite Galois groups, and hence the $H^i(\mathrm{Gal}(K/k)/H, A^H)$ are the usual Galois cohomology groups from Lecture 28, without having to worry about any continuity condition.

(v) It thus follows from Hilbert's Theorem 90 for finite Galois extensions that the same also applies to infinite Galois extensions, so that $H^1(\mathrm{Gal}(K/k), K^\times) = 0$ (as long as we mean H^1 in the sense being discussed, taking the topology of $\mathrm{Gal}(K/k)$ into account).

(vi) When $i = 1$, one can analogously talk of pointed sets $H^1(\mathrm{Gal}(K/k), M)$ for possibly nonabelian groups M on which the possibly infinite group $\mathrm{Gal}(K/k)$ acts continuously. In this context, we still have Hilbert's Theorem 90, saying that $H^1(\mathrm{Gal}(K/k), GL_n(K))$ is trivial for each n.

29.6. **Transcendental extensions.**

Definition 29.14. (i) A field extension K/k is said to be transcendental if it is not algebraic. Note that K/k is transcendental if and only if there exists $\alpha \in K$ such that the k-algebra homomorphism $k[x] \to K$ sending x to α is injective, in which case the subfield $k(\alpha) \subset K$ generated by k and α is isomorphic to the field $k(x)$ of rational functions in one variable over k. Such an α is said to be transcendental over k.

(ii) Let K/k be a field extension. A subset $S \subset K$ is said to be algebraically independent over k if the elements of S do not satisfy any nontrivial polynomial relation with coefficients in k; more precisely, if every k-algebra homomorphism $\varphi : k[x_1, \ldots, x_n] \to K$ with $\varphi(x_i) \in S$ for all $1 \leqslant i \leqslant n$ is injective.

(iii) A subset $S \subset K$ which is algebraically independent over k, and is maximal with respect to inclusion, is called a transcendence base or a transcendence basis for K over k. Note that $S \subset K$ is a transcendence basis if and only if the following two conditions hold: S is algebraically independent over k, and K is algebraic over $k(S)$ (where $k(S)$ stands for the subfied of K generated by k and S; note that it is isomorphic to the quotient field $k(x_s \mid s \in S)$ of a polynomial ring $k[x_s \mid s \in S]$ in variables indexed by S).

Theorem 29.15. *Let K/k be a field extension. Then any two transcendence bases for K over k have the same cardinality.*

Sketch of the proof in the special case where K has a finite transcendence basis. We follow Serge Lang's book. Suppose K has a finite transcendence basis $\{x_1, \ldots, x_m\}$. It is enough to show that if $w_1, \ldots, w_n \in K$ are algebraically independent over k, then $n \leqslant m$ (this suffices by symmetry, which will then give the inequality $m \leqslant n$ whenever $w_1, \ldots, w_n$ is a transcendence basis).

Since $w_1, x_1, \ldots, x_m$ are algebraically dependent, we may write $f_1(w_1, x_1, \ldots, x_m) = 0$ for some irreducible polynomial $f_1 \in k[Y_0, \ldots, Y_m]$ over k in $m+1$ variables. Without loss of generality, we may assume that "Y_1 occurs in f_1", i.e., that we can write

$$0 = f_1(w_1, x_1, \ldots, x_m) = \sum_j g_j(w_1, x_2, \ldots, x_m) x_1^j,$$

with $g_N \neq 0$ for some $N \geqslant 1$ (as a polynomial in m variables).

Note that no irreducible factor of g_N vanishes on $(w_1, x_2, \ldots, x_m)$: otherwise w_1 is a root of two distinct irreducible polynomials over $k(x_1, \ldots, x_m)$.

Therefore, x_1 is algebraic over $w_1, x_2, \ldots, x_m$, and $w_1, x_2, \ldots, x_m$ are algebraically independent over k (otherwise $x_1, \ldots, x_m$ would be algebraically dependent over k).

Now one repeats this argument, assuming that $w_1, \ldots, w_r, x_{r+1}, \ldots, x_m$ are algebraically independent and that $x_1, \ldots, x_r$ are algebraic over $w_1, \ldots, w_r, x_{r+1}, \ldots, x_m$, and then shows (after renumbering the x_i if necessary) that $w_1, \ldots, w_{r+1}, x_{r+2}, \ldots, x_m$ is algebraically independent, and that K is algebraic over $k(w_1, \ldots, w_{r+1}, x_{r+2}, \ldots, x_m)$. Therefore, one can proceed by induction. □

Exercise 29.16. Let K/k be a field extension. Show that any set of elements in K that are algebraically independent over k can be completed into a transcendence basis for K over k. Moreover, given any set $\Gamma \subset K$ of generators for K/k (i.e., the minimal subfield $k(\Gamma) \subset K$ containing Γ is K), show that the extra elements in this transcendence basis can be chosen to be from Γ.

Definition 29.17. If K/k is a field extension, the transcendence degree of K/k, sometimes denoted $\mathrm{trdeg}_{K/k}$, is the cardinality of some, and hence by Theorem 29.15 any, transcendence basis for K over k.

Exercise 29.18. Read up about the notion of separable extensions in the context of transcendental extensions.

29.7. **Kummer theory – the simplest case.** We will follow Serge Lang's book.

Notation 29.19. (i) In this subsection, let k be a field, and m a positive integer, such that:
- $(\mathrm{char}\, k, m) = 1$ (considered automatically true if $\mathrm{char}\, k = 0$); and
- $\mu_m(k) = \mu_m(\bar{k}) \subset k^\times$, where we write $\mu_m(k)$ and $\mu_m(\bar{k})$ for the sets of m-th roots of unity in k and some algebraic closure $\bar{k}$ of k. In other words, given that $(\mathrm{char}\, k, m) = 1$, we are assuming that k contains m distinct m-th roots of unity; call their set μ_m.

(ii) Fix an algebraic closure $k \hookrightarrow \bar{k}$; in this subsection, "an extension of k" will refer to "an extension of k in $\bar{k}$".

Kummer theory answers the question: what are the finite abelian extensions of K/k (by convention, inside $\bar{k}$) [88] of exponent m?

To state the answer, we will use the following notation:

Notation 29.20. (i) We will write $(k^\times)^m$ (resp., k^m) for $\{a^m \mid a \in k^\times\}$ (resp., $\{a^m \mid a \in k\}$).

(ii) For $a \in k \backslash k^m = k^\times \backslash (k^\times)^m$, $k(a^{1/m}/k)$ will denote the extension K/k in $\bar{k}$ obtained by adjoining a choice of an m-th root $a^{1/m}$ of a in $\bar{k}$: note that while $a^{1/m}$ itself is only well-defined up to multiplication by an element of $\mu_m(\bar{k}) = \mu_m(k)$, $k(a^{1/m})$ is well-defined because of our hypothesis that $\mu_m(\bar{k}) \subset k$. Thus, $k(a^{1/m})$ is also the extension of k in $\bar{k}$ obtained by adjoining all the m-th roots of a in $\bar{k}$.

(iii) Given $B \subset k$, $k(B^{1/m})$ will denote the extension K/k obtained by adjoining to k all the m-th roots of the elements of B in $\bar{k}$.

Then the answer give by Kummer theory in our easy situation is:

Theorem 29.21. *Assume Notation 29.19, so* $(\mathrm{char}\, k, m) = 1$ *and* $\mu_m(\bar{k}) = \mu_m(k)$. *Then there exists a bijection*

$$\{\textit{Subgroups of } k^\times \textit{ containing } (k^\times)^m\} \to \{\textit{Abelian extensions of } k \textit{ of exponent } m\},$$

sending B *to* $K_B := k(B^{1/m})$.

Note that the following lemma is an analogue of the Artin-Schreier theorem discussed in Lecture 28 (Theorem 28.23). Just as the proof of the Artin-Schreier theorem used the additive form of Hilbert's theorem 90, the following lemma will use the multiplicative form. This lemma will be a sort of induction step in the proof of Theorem 29.21.

[88] Note that multiple such extensions can be isomorphic in $f\acute{e}t_k$.

Lemma 29.22. *(i) Let K/k be a finite cyclic extension of degree n. Then there exists $\alpha \in K$ such that $K = k(\alpha)$, and such that α is an n-th root of some $a \in k$.*

(ii) Let $a \in k$. Then $k(a^{1/m})$ is Galois, with Galois group isomorphic to $\mathbb{Z}/d\mathbb{Z}$ for some $d|m$: in fact sending $\sigma \in \mathrm{Gal}(k(a^{1/m})/k)$ to $\sigma(a^{1/m})/a^{1/m}$ gives an injective homomorphism of groups $\mathrm{Gal}(k(a^{1/m})/k) \hookrightarrow \mu_m$ (and hence has cyclic image of order some $d|n$).

Proof. The proof of (i), which uses Hilbert's Theorem 90 analogously to how the proof of the corresponding assertion of the Artin-Schreier theorem used the additive form of Hilbert's Theorem 90, is part of HW 14.

$k(a^{1/m})/k$ is Galois, because it is separable and normal – the former because $(m, \mathrm{char}\, k) = 1$), and the latter because it is clearly it contains the set $a^{1/m}\mu_m(k) = a^{1/m}\mu_m(\bar{k})$ of all roots of $x^m - a$, and is hence the splitting field of $x^m - a$. The rest of (ii) is immediate, but let us remark that $\sigma(a^{1/m})/a^{1/m}$ is independent of the choice of $a^{1/m}$ because $\mu_m(\bar{k}) \subset k$. □

Proposition 29.23. *Assume Notation 29.19. Let $(k^\times)^m \subset B \subset k^\times$ be a subgroup. Let $K_B = k(B^{1/m})$.*

(i) $K_B = k(B^{1/m})$ is a Galois extension of k.

(ii) Set $G = \mathrm{Gal}(K_B/k)$. Consider the bilinear map

$$G \times B \to \mu_m, \qquad (\sigma, a) \mapsto \frac{\sigma(a^{1/m})}{a^{1/m}}.$$

The kernel of this map on the left is $\{1\} \subset G$, and the kernel of this map on the right is $(k^\times)^m \subset B$.

(iii) K_B/k is finite if and only if $[B : (k^\times)^m]$ is finite, in which case we have $B/(k^\times)^m \cong \mathrm{Hom}(G, \mu_m)$ and $[K_B : k] = [B : (k^\times)^m]$.

Proof. (i) is easy, just like in Lemma 29.22(ii).

Now we come to (ii). If $\sigma \in G$ belongs to the kernel on the left, then $\sigma(a^{1/m}) = a^{1/m}$ for all $a \in B$, so σ is the identity on K_B. Thus, the kernel on the left is $\{1\}$. If $a \in B$ is such that $\langle \sigma, a \rangle = 1$ for all $\sigma \in G$, then (any choice of) $a^{1/m}$ is fixed by all $\sigma \in G = \mathrm{Gal}(K_B/k)$, so $k(a^{1/m}) = k$, so that $a \in (k^\times)^m$. This proves (ii).

(iii) is then an easy exercise using (ii) and basic properties of pairings between finite groups. □

Proof of Theorem 29.21, somewhat tersely written. First we prove the injectivity of $B \mapsto K_B$. It is easy to see that if $B_1 \subset B_2$ then $K_{B_1} \subset K_{B_2}$.

Conversely, suppose $K_{B_1} \subset K_{B_2}$, but there exists $b_1 \in B_1 \backslash B_2$. Then $k(b_1^{1/m}) \subset K_{B_1} \subset K_{B_2}$ is contained in a finitely generated subextension of K_{B_2}. This allows us to assume without loss of generality that $[B_2 : (k^\times)^m]$ is finite (replace B_1 by the image of $\langle b_1 \rangle$ in $k^\times/(k^\times)^m$, and B_2 by a suitable smaller group while ensuring that we still have $k(b_1^{1/m}) \subset K_{B_2}$). Let B_3 be the subgroup of $k^\times$ generated by B_2 and b. Then $K_{B_2} = K_{B_3}$, so by Proposition

29.23(iii), we have $\#(B_2/(k^\times)^m) = \#(B_3/(k^\times)^m)$, so $B_2 = B_3$, a contradiction to $b_1 \notin B_2$. This gives the injectivity of $B \mapsto K_B$.

Now we come to the surjectivity, so let K be an abelian extension of k of exponent m. It is enough to see that K is a compositum of extensions of the form $k(a^{1/m})$ (since then we can let B be the subgroup of $k^\times$ generated by $(k^\times)^m$ and all those a). Equivalently, it is enough to see that any finite (necessarily Galois, by the abelianness of K/k) subextension of k is a compositum of extensions of the form $k(a^{1/m})$.

Being abelian of exponent m, it is easy to see that any such finite extension is a compositum of cyclic extensions of exponent m: this is an easy argument involving the structure theorem for abelian groups. Thus, we are done since, by Lemma 29.22, any cyclic extension of k of degree m is obtained by adjoining an m-th root of some $a \in k$. □

29.8. **Appendix – The Brauer group and Galois cohomology.** This subsection was not discussed in the lecture, probably not even alluded to, and hence is optional. Recall central simple algebras and Brauer groups from Lecture 19. In this subsection, we wish to study how to relate Brauer groups to Galois cohomology. Specifically, given a finite Galois extension K/k, we would like to get an isomorphism between a subgroup $Br(K/k) \subset Br(k)$, and $H^2(\mathrm{Gal}(K/k), K^\times)$. We will only sketch a proof of a bijection (with no indication that it is an isomorphism of groups), and that too modulo some black-boxes.

Definition 29.24. Let $Br(K/k)$ denote the subset of $Br(K)$ represented by central simple algebras A/k with the property that A splits over K, i.e., $A \otimes_k K \cong M_n(K)$ for some n.

Blackbox: We will with very little further mention assume the following as a blackbox: A central simple algebra A over k represents an element of $Br(K/k)$ if and only if there exists an A' in its equivalence class such that the k-algebra K embeds into A' as a maximal commutative subfield. For a proof, see Lemma 16.3 of

`https://ocw.mit.edu/courses/18-706-noncommutative-algebra-spring-2023/mit18_706_s23_lec16.pdf`

for a proof. Note that A' as above is unique by dimension considerations.

Recall from Theorem 29.8 that $H^2(\mathrm{Gal}(K/k), K^\times)$ is in a certain bijection with the set of isomorphism classes of extensions

$$1 \to K^\times \to N \to \mathrm{Gal}(K/k) \to 1. \tag{146}$$

Thus, we need in particular a bijection from isomorphism classes of extensions (146), and $Br(K/k)$. For this, we make:

Construction 29.25. Let K/k be a finite Galois extension. Given a central simple algebra A over k into which the k-algebra K embeds as a maximal subfield, fix such an embedding $K \hookrightarrow A$, and let N be the normalizer of $K^\times$ in $A^\times$. Therefore, N defines, by conjugation, an action of $N/K^\times$ on K, and hence a homomorphism $N/K^\times \to \mathrm{Gal}(K/k)$. This map is

injective since $K \subset A$ is a maximal commutative subfield, and it is easy to see from the Skolem-Noether theorem that this map is surjective (if $\sigma \in \mathrm{Gal}(K/k)$, then $(K \hookrightarrow A) \circ \sigma$ is another k-algebra embedding of K into A). Thus, we get an extension

$$1 \to K^\times \to N \to \mathrm{Gal}(K/k) \to 1 \tag{147}$$

of the abelian group $K^\times$ by the group $\mathrm{Gal}(K/k)$. It is immediate that the action of $\mathrm{Gal}(K/k)$ on $K^\times$ associated to this extension is just the usual action of $\mathrm{Gal}(K/k)$ on $K^\times$.

Using the Skolem-Noether theorem, it is easy to check that the isomorphism class of the extension (147) is independent of the choice of the embedding $K \hookrightarrow A$ (this is an important point). Note also from the blackbox above that A is the unique central simple algebra in the equivalence class associated to its image in $Br(K/k)$ with K as a maximal commutative subfield.

The theorem I wish to discuss is:

Theorem 29.26. *Let K/k be a finite Galois extension. Then there is a unique isomorphism of groups $Br(K/k) \to H^2(\mathrm{Gal}(K/k), K^\times)$, which sends the image of each central simple algebra A over k having the k-algebra K as a maximal commutative subfield (see the blackbox above), to the element of $H^2(\mathrm{Gal}(K/k), K^\times)$ that corresponds, by the bijection of Theorem29.8, to the isomorphism class of the extension of $\mathrm{Gal}(K/k)$ by $K^\times$ associated to A in Construction 29.25. It has a suitably functorial dependence on K.*

Remark 29.27. Suppose K/k is infinite Galois. Letting $H^2(\mathrm{Gal}(K/k), K^\times)$ be as in Remark 29.13, i.e., Galois cohomology with continuous cocycles, and using some basic facts about group cohomology, it is easy to deduce that the assertion $Br(K/k) \cong H^2(\mathrm{Gal}(K/k), K^\times)$ continues to hold in this case, so in particular we have $Br(k) \cong H^2(\mathrm{Gal}(k^s/k), (k^s)^\times)$, where $k \hookrightarrow k^s$ is a separable closure.

Some ideas involved in a proof of Theorem 29.26. Please be careful of several potential inaccuracies, even serious ones, in the following: I could have been careless; I think I did not follow any standard reference except for the description of the construction of A_c.

Already, Construction 29.25 (including the observation at the end that it is well-defined), together with Theorem 29.8, gives us a map $Br(K/k) \to H^2(\mathrm{Gal}(K/k), K^\times)$ – here we have used the blackbox above.

Let us describe how to attach a central simple algebra A_c to a 2-cocycle $c := (c_{\sigma,\tau})_{\sigma,\tau}$ of $G := \mathrm{Gal}(K/k)$ in $K^\times$. For simplicity we will assume c to be normalized, so that $c_{1,\sigma} = c_{\sigma,1} = 1 \in K^\times$ for all σ. To c we attach:

$$A_c = \Big\{ \sum_{\sigma \in G} a_\sigma [\sigma] \mid a_\sigma \in K \,\forall\, \sigma \Big\},$$

and define a multiplication on A_c by imposing the relations $[\sigma] \cdot x = \sigma(x) \cdot [\sigma]$ and $[\sigma][\tau] = c_{\sigma,\tau}[\sigma\tau]$, for all $\sigma, \tau \in \mathrm{Gal}(K/k)$ and $x \in K$.

Just as with group extensions, one deduces associativity from the 2-cocycle condition. It is now easy to see that A_c is a ring that is a k-vector space of dimension $[K:k]^2$. Via the inclusion $a \mapsto a[1]$, A_c is a k-algebra (this uses that the cocycle c is normalized). $a \mapsto a[1]$ also gives us a k-algebra embedding $K \hookrightarrow A_c$, which we will think of as an inclusion for the rest of this proof.

Let us verify that A_c is a central simple algebra. Since conjugation by $\alpha \in K^\times$ sends $\sum a_\sigma[\sigma]$ to $\sum a_\sigma \cdot (\sigma(\alpha)\alpha^{-1}) \cdot [\sigma]$, it follows that the centralizer of $K^\times$ in A_c equals $K = K \cdot [1] \subset A_c$. Therefore, the center of A_c is contained the subset of K fixed by each $[\sigma]$, namely, $k \subset K$. Thus, A_c is central.

Let us show that A_c is simple. Suppose $I \subset A_c$ is a two-sided ideal. Again, consider conjugation by $K^\times$. On the basis element $[\sigma]$ of A_c as a K-vector space (where K acts on A_c by left multiplication), conjugation by $\alpha \in K^\times$ acts as multiplication by $\alpha\sigma(\alpha^{-1})$. Thus, this conjugation action of $K^\times$ on A_c is simultaneously diagonalizable. Since this action preserves I, it follows that I is a direct sum of simultaneous eigenspaces for this action (a submodule of a semisimple module is semisimple). Thus, I is a linear span of some of the $[\sigma]$ (note that each $K \cdot [\sigma]$ is the whole of a simultaneous eigenspace). If I is nonzero, then it contains some $[\sigma]$, from which, multiplying by other $[\tau]$ and some nonzero scalars, it follows that I contains every $[\sigma]$ and hence equals A_c. Thus, we have shown that A_c is simple as well, so it is central simple over k. Since we have a k-algebra embedding $K \hookrightarrow A_c$, A_c represents an element of $Br(K/k)$ (use the blackbox above).

Further, it is easy to see that if $(c'_{\sigma,\tau})_{\sigma,\tau}$ is cohomologous to $(c_{\sigma,\tau})_{\sigma,\tau}$, with $c_{\sigma,\tau} - c'_{\sigma,\tau} = \sigma_1(b_{\sigma_2}) - b_{\sigma_1\sigma_2} + b_{\sigma_1}$, then mapping $\sum a_\sigma[\sigma]$ to $\sum a_\sigma b_\sigma[\sigma]$ defines an isomorphism of algebras $A_c \to A_{c'}$ (we are being slightly vague here: this is true if c and c' are normalized; otherwise one can use this to realize A_c and A'_c as k-algebras, and then the same statement will hold). Thus, sending c to A_c induces a well-defined set-theoretic map $H^2(\mathrm{Gal}(K/k), K^\times) \to Br(K/k)$. To show that this map is a bijection, it remains to show that $H^2(\mathrm{Gal}(K/k), K^\times) \to Br(K/k) \to H^2(\mathrm{Gal}(K/k), K^\times)$ and $Br(K/k) \to H^2(\mathrm{Gal}(K/k), K^\times) \to Br(K/k)$ are the identity maps.

What do we get when we apply Construction 29.25 to A_c, or rather what is the element of $H^2(G, K^\times)$ attached by Theorem 29.8 to the resulting extension? The normalizer of $K^\times \subset K \subset A_c$ in $A_c^\times$ contains each $[\sigma] \in A_c$, conjugation by which is readily verified to send $\alpha \in K^\times$ to $\sigma(\alpha)$. Thus, it is automatic from (147) that $K^\times$ together with the $[\sigma]$ generate N. Using $\sigma \mapsto [\sigma]$ in place of the section s in the proof of Theorem 29.8, we find that the element of $H^2(\mathrm{Gal}(K/k), K^\times)$ associated to the extension given by applying Construction 29.25 to A_c, is represented by the two-cocycle

$$(\sigma, \tau) \mapsto [\sigma][\tau][\sigma\tau]^{-1} = c_{\sigma,\tau} \in K^\times.$$

This shows that $H^2(\mathrm{Gal}(K/k), K^\times) \to Br(K/k) \to H^2(\mathrm{Gal}(K/k), K^\times)$ is the identity map.

Now let us show that $Br(K/k) \to H^2(\mathrm{Gal}(K/k), K^\times) \to Br(K/k)$ is the identity map. We use the blackbox above to start with a representative A for a given element of $Br(K/k)$,

that contains the k-algebra K as a maximal commutative subfield. We look at (147) in Construction 29.25, and assume notation from there. We let $[\sigma] \in N$ to be any preimage of $\sigma \in \mathrm{Gal}(K/k)$, except that we require $[1] = 1$. Then it follows from Construction 29.25 that $[\sigma][\tau] = c_{\sigma,\tau}[\sigma\tau]$ for some 2-cocycle $(c_{\sigma,\tau})_{\sigma,\tau}$ representing the extension given by that construction. Since $[1] = 1$, it is easy to see that the 2-cocycle $(c_{\sigma,\tau})_{\sigma,\tau}$ is normalized. Given the construction of $H^2(\mathrm{Gal}(K/k), K^\times) \to Br(K/k)$ via $c \mapsto A_c$, it now suffices to show that as a K-vector space (K operating on A by left multiplication), A has the $[\sigma]$ as a basis.

By dimension considerations, and the fact that $[K : k] = \mathrm{Gal}(K/k)$, this follows if we show that the $[\sigma]$ are K-linearly independent. But this follows from the fact that conjugation by $\alpha \in K^\times$ acts on $K \cdot [\sigma]$ by multiplication by $\alpha\sigma(\alpha^{-1})$, and the simultaneous eigencharacters $\alpha \mapsto \alpha \cdot \sigma(\alpha^{-1})$ are pairwise distinct.

□