MATH 210A: Modern Algebra I

Lecturer: Professor Richard Taylor

Notes by: Andrew Lin

Autumn 2022

Introduction

MATH 210A is a course that aims to cover the material covered on the graduate qualifying exam. It's meant to be a collection of useful techniques, but it won't culminate in any big theorem, so it's not a course we take for relaxation – we should be dedicated to it. This course will focus on mathematical structures (algebraic objects and maps between them), rather than how to combine their elements. So in some sense, the level of abstraction will be "one level up."

This class will cover (only) **commutative** algebra. The first three weeks of the class will discuss rings – constructions of rings, products and coproducts, quotients, localizations, completions, and polynomial rings. The Noetherian property and prime ideals will make an appearance, and there will be some discussion of unique factorization domains. The next week will cover category theory, which abstracts the ideas of structures and maps in the same way that groups generalize collections of symmetries. After that, the next three weeks will discuss modules over rings (basic properties, multilinear algebra, Nakayama's lemma, and the structure of modules over a principal ideal domain and its applications to linear algebra). The final three weeks will cover an introduction to homological algebra (abelian categories, injectives, projectives, exact sequences, derived functors, Ext and Tor), and sheaves will be covered as an example (though they are not officially on the qualifying exam syllabus).

Quadratic algebra and bilinear forms are topics that we should be covering for the qualifying exam as well, but we won't have time to cover that in class, so there is a handout on Canvas about this material which we should review for the exam. (In other words, it's examinable on the qual but not for this class.) On that note, this class will be fast-paced, and we'll have to do the work to keep up.

There won't be much prerequisite knowledge, but we will assume previous experience with abstract mathematics. There are lots of good textbooks on this material – Dummit and Foote is relatively elementary compared to what we'll cover, but it's a good one to learn from and sufficient for the qualifying exam. Jacobson's book is the official recommended text, and it's a great reference to have. But it's difficult to learn from (at least for this course) because that book emphasizes noncommutative algebra instead for more efficiency, and the intuitions are different in commutative and noncommutative algebra. Finally, previous iterations of this class have used Aluffi's book, which follows a similar path of emphasizing categorical underpinnings.

Mathematics has to be learned by doing, so exercises are crucial for the course – we will have weekly assignments, and the first one is on Canvas and due at the start of class last Monday. Professor Taylor can be reached at rltaylor@stanford.edu, and the course assistant (grader) for this class is Lie Qian (lqian@stanford.edu). Office hours will provisionally be set at Thursday 1:30-2:30pm (in person) for Professor Taylor and Tuesday 8-9am (Zoom) and Friday 4-6pm (in person) for Lie. We're encouraged to work on homework together but to work on the problems ourselves before doing so, or else we might be a passive participant and not really realize the difficulties and important parts of the problems.

The class will also have weekly quizzes – Professor Taylor sees them as better than midterms because we have to keep on top of the material every week this way. They'll be open book and online, and we have a 24-hour window on Thursday to complete them (but we have only 65 minutes to upload our solutions after we download the quiz). Finally, there will be a traditional closed-book final exam during exam period. (No collaboration is allowed on the quizzes or final exam.)

We're encouraged to ask questions during lecture – the point of an in-person instruction is to be interactive, so we should speak up whenever we have a question or comment. And attending office hours is good too – mathematicians often get stuck on what turns out to be an easy point in the end. It's often just about a change of perspective and can get sorted out if we ask others for help!

For this quarter, the university requires students to wear masks when taking classes for credit (so we should be aware of that). Finally, anyone who needs accommodations should let Professor Taylor know as soon as possible.

1 September 26, 2022

Many of us have probably seen rings in some sense before, so this will be covered quickly. There will be notes about rings on Canvas if we want to check our understanding in any way.

Definition 1

A **ring** is a set R containing two distinguished elements $0, 1$ and two binary operations $+$ (addition) and $\cdot$ (multiplication), such that (1) $(R, +, 0)$ is an abelian group (addition is associative and commutative, 0 is the identity element, and every r has an additive inverse $-r$), (2) $\cdot$ is associative, commutative, and has identity element 1, and (3) the distributive property $r(s+t) = rs + rt$ holds.

We will only consider commutative rings with a multiplicative identity, though some textbooks may consider rings in more generality. Associativity means that brackets don't matter, so we can write $r+s+t$ or rst without needing to worry about the order in which operations are performed.

Example 2

The trivial ring containing only the element $\{0\}$ (so that $0 = 1$) satisfies all of the axioms. More interesting examples of rings that we may have seen include $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, and $\mathbb{Z}/n\mathbb{Z}$ (which will also be denoted $\mathbb{Z}/(n)$).

Example 3

The set of continuous functions $C([0,1]) = \{f : [0,1] \to \mathbb{C} : f \text{ continuous}\}$ forms a ring with addition and multiplication defined pointwise (so that $(f+g)(t) = f(t) + g(t)$ and $(fg)t = f(t)g(t)$). The additive and multiplicative identity are then the 0 and 1 functions, respectively.

Example 4

The set $\{(a, b) \in \mathbb{Z}^2 : a \equiv b \bmod 3\}$ with component-wise addition and multiplication also forms a ring.

The following simple facts are good exercises if we don't see how they work immediately:

- For any $r \in R$, we have $-r = (-1) \cdot r$.

- For any $r \in R$, $0 \cdot r = 0$.
- We only have $0 = 1$ if R is the trivial ring.

Definition 5

Even though multiplication is not required to have inverses, we can define the **group of units of R**

$$R^\times = \{r \in R : \exists s \in R \text{ with } rs = 1\}.$$

(We can check that multiplicative inverses are unique if they exist.)

For example, the group of units in $\mathbb{Z}^\times$ is $\{\pm 1\}$, and the group of units in $\mathbb{Q}^\times$ is $\mathbb{Q} \setminus \{0\}$. (And the group of units in $C([0,1])$ is the set of nowhere zero functions.)

Definition 6

An element $r \in R$ is **nilpotent** if $r^n = 0$ for some integer $n > 0$.

For example, 0 is always nilpotent, and $2^2 = 0$ in $\mathbb{Z}/4\mathbb{Z}$ so 2 is nilpotent in $\mathbb{Z}/4\mathbb{Z}$.

Definition 7

An element $r \in R$ is a **zero-divisor** if there is some nonzero $s \in R$ such that $rs = 0$.

All nilpotent elements are zero divisors, and the function

$$f(x) = \begin{cases} x - \frac{1}{2} & x \in \left[0, \frac{1}{2}\right] \\ 0 & x \in \left[\frac{1}{2}, 1\right] \end{cases}$$

is a zero-divisor because it can be multiplied by the function $g(x) = \begin{cases} 0 & x \in \left[0, \frac{1}{2}\right] \\ x - \frac{1}{2} & x \in \left[\frac{1}{2}, 1\right] \end{cases}$.

Definition 8

A ring is **reduced** if 0 is the only nilpotent element, an **integral domain** (ID) if 0 is the only zero divisor, and a **field** if $R^\times = R \setminus \{0\}$.

In particular, any field is an integral domain because all nonzero elements have inverses, and any integral domain is reduced. For example, $\mathbb{Z}$ is an integral domain but not a field, $\mathbb{Q}$ is a field, $\mathbb{Z}/6\mathbb{Z}$ is reduced, and $\mathbb{Z}/4\mathbb{Z}$ isn't even reduced. The trivial ring is a special case – it is not an integral domain or a field because of technicalities with zero itself, but it is reduced. (And it's important that the definitions are set up so that this is the case.)

Whenever we have a structure, we'll want to consider the maps that preserve that structure:

Definition 9

A map $\phi : R \to S$ is a **ring (homo)morphism** if it preserves the structure of addition and multiplication – in other words, $\phi(0) = 0, \phi(1) = 1$, and $\phi(r + s) = \phi(r) + \phi(s)$ and $\phi(rs) = \phi(r)\phi(s)$ for $r, s \in R$.

Some books may not make the requirement that $\phi(1) = 1$ (in which case some "categorical properties" break down), but it's the right thing to do when we're studying commutative algebra. And those conditions above are enough to guarantee that negatives are preserved as well: $\phi(-r) = -\phi(r)$.

Example 10

The inclusion $Z \hookrightarrow \mathbb{C}$ and the evaluation map $C[0,1] \to \mathbb{C}$ sending f to $f(\frac{1}{2})$ are both ring homomorphisms.

Example 11

The map $R \to \{0\}$ sending everything to 0 is a ring homomorphism for any R ("any ring has a unique homomorphism to the trivial ring"), but the inclusion $\{0\} \hookrightarrow \mathbb{Z}$ is not a ring homomorphism because 1 is not sent to 1. On the flip side, "there is a unique morphism from $\mathbb{Z}$ to any other ring," since 0 must be sent to 0 and n must be sent to $1 + 1 + \cdots + 1$ (added together n times) for any $n > 0$, while $-n$ must be sent to $-(1 + 1 + \cdots + 1)$. (It's a tedious exercise to check that this does work.)

Definition 12

A subset $R \subset S$ is a **subring** if $0, 1 \in R$ and R is closed under addition, multiplication, and negatives. (In other words, R itself should be a ring with respect to those operations.)

Our early discussions of rings will focus on constructing new rings, and the easiest is the following:

Definition 13

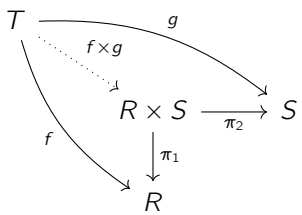
For any rings R, S, the **product ring** $R \times S = \{(r, s) : r \in R, s \in S\}$ is defined by having addition and multiplication performed component-wise (so that $0 = (0, 0)$ and $1 = (1, 1)$).

This product ring $R \times S$ comes with natural projections $\pi_1 : R \times S \to R$ (sending $(r, s) \to r$) and $\pi_2 : R \times S \to S$ (sending $(r, s) \to s$), both of which are ring morphisms. But this doesn't work the other way around: mapping $Z \to \mathbb{Z} \times \mathbb{Z}$ sending n to $(n, 0)$ is not a ring morphism because it doesn't send 1 to 1. (Other algebraic objects have "maps in" and "maps out," but this is not the case for rings.)

Lemma 14

Let T be any ring, and suppose that $f : T \to R$ and $g : T \to S$ are morphisms of rings. Then there exists a unique map $f \times g : T \to R \times S$ such that $\pi_1 \circ (f \times g) = f$ and $\pi_2 \circ (f \times g) = g$.

This is often written in a diagram as shown below:



Basically, the map shown with the dotted arrow must uniquely exist, and the diagram **commutes** because we get the same answer no matter which way we follow the arrows around from one point to another. And constructing the

map itself is easy here: we must have $(f \times g)(t) = (f(t), g(t))$, and this does indeed define a morphism. And this lemma "uniquely characterizes" the product ring $R \times S$ – any other ring satisfying this **universal property** must be isomorphic to the product ring. But we'll discuss this fact more next time.

2 September 28, 2022

As a reminder, homework should be submitted in class on paper (though we can type up our solutions and print them out).

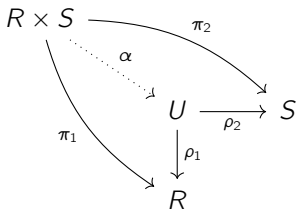
Last lecture, we started discussing constructions of rings, starting with the product ring $R \times S = \{(r, s) : r \in R, s \in S\}$ defined by component-wise addition and multiplication. In particular, we mentioned that such a ring always comes with projection maps $\pi_1 : R \times S \to R$ and $\pi_2 : R \times S \to S$, and that for any ring T with maps $f : T \to R$ and $g : T \to S$, there is a natural map $f \times g : T \to R \times S$ completing the "commutative diagram" yielding $f = \pi_1 \circ (f \times g)$ and $g = \pi_2 \circ (f \times g)$.

It's not too hard to show why this universal property holds, but we'll see many arguments of this form in the rest of the course, so we'll go through the argument here.

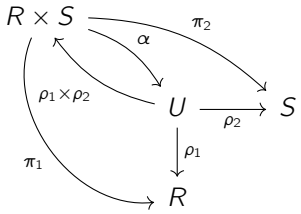
Proposition 15

Suppose there were another ring U with morphisms $\rho_1 : U \to R$ and $\rho_2 : U \to S$ with the universal property. Then $U \cong R \times S$.

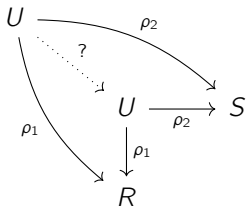
Proof. Apply the universal property of U with $T = R \times S$. Then we have the commutative diagram shown below, yielding a unique map $\alpha : R \times S \to U$.



But we can also swap the roles of $R \times S$ and U, applying the universal property of $R \times S$ with $T = U$, to get a map $\rho_1 \times \rho_2 : U \to R \times S$.



Finally, if U appears in both spots, we see that the diagram commutes whether the dotted map is the identity map (since $\rho_1 = \rho_1 \circ \mathrm{Id}_U$ and so on) or $\alpha \circ (\rho_1 \times \rho_2)$ (by following the arrows in the previous diagrams, we see that $\rho_2 = \pi_2 \circ (\rho_1 \times \rho_2) = \rho_2 \circ \alpha \circ (\rho_1 \times \rho_2)$ so the top loop commutes, and similarly the bottom loop commutes).



Thus $\mathrm{Id}_U = \alpha \circ (\rho_1 \times \rho_2)$ (because the map must be unique), and similarly we find that $\mathrm{Id}_{R\times S} = (\rho_1 \times \rho_2) \circ \alpha$, which yields an isomorphism betewen U and $R \times S$. □

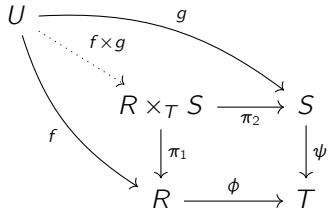
Two rings can be isomorphic in various ways, and the one we've described is in some way "natural:" the isomorphism α we described above is unique if we require that $\rho_2 \circ \alpha = \pi_2$ and $\rho_1 \circ \alpha = \pi_1$. And later, when we discuss things like tensor products, we'll see that this type of argument works more easily than trying to wrestle with the constructions directly.

Extending our definition of the product, we can take any index set I and construct the product ring $\prod_{i\in I} R_i$. Then everything we've said generalizes whether I is finite or infinite.

Definition 16

Suppose R, S, T are rings, and we have ring morphisms $\phi : R \to T$ and $\psi : S \to T$. The **relative product** $R \times_{\phi,T,\psi} S$ (often denoted $R \times_T S$) is the set $\{(r, s) \in R \times S : \phi(r) = \psi(s)\}$ as a subring of $R \times S$.

This relative product also has a universal property: we still have projection maps $\pi_1 : R \times_T S \to R$ and $\pi_2 : R \times_T S \to S$ defined by restriction. Then if $\phi \circ f = \psi \circ g$ in the diagram below, then there exists a unique map $f \times g : U \to R \times_T S$ which makes the diagram commute.



Example 17

We mentioned the ring $\{(a, b) \in \mathbb{Z}^2 : a \equiv b \bmod 3\}$ last time, which can also be represented as the relative product $\mathbb{Z} \times_{\mathbb{Z}/3\mathbb{Z}} \mathbb{Z}$.

We'll next discuss **polynomial rings**:

Definition 18

Let x_i be a set of (commuting) indeterminates (where $i \in I$ are part of some index set). We may form **monomials** $x_{i_1}^{n_1} \cdots x_{i_r}^{n_r}$ as finite products of the x_i. Then for any ring R, the **polynomial ring** $R[x_i]$ is the set of formal finite sums of elements of R multiplied by monomials, with addition and multiplication defined in the usual way, and the **formal power series ring** $R[[x_i]]$ is the same but with ∞-formal sums allowed.

There is an embedding $R \hookrightarrow R[x_i]$ sending r to r times the trivial monomial, and $R[x_i] \hookrightarrow R[[x_i]]$ because any polynomial is a power series. Also, we can check that $(R[x])[y] = R[x, y]$ and so on.

Remark 19. *It's okay to have uncountably many variables in all of these definitions, but we can still define (for example) multiplication of formal power series because there are only finitely many (pairs of) terms that can contribute to any particular monomial in the product.*

Definition 20

Let $f \in R[x]$ be a polynomial. The **degree** of f is $\deg(f) = -\infty$ if $f = 0$, and otherwise if $f = a_0 + a_1 x + \cdots + a_d x^d$ with $a_d \neq 0$, then $\deg(f) = d$. We call f **monic** if $a_d = 1$.

Lemma 21

If R is an integral domain, then $\deg(fg) = \deg(f)\deg(g)$ for all $f, g \in R[x]$.

Proof. If the highest-degree terms of f and g are $a_d x^d$ and $b_e x^e$, respectively, then $a_d b_e \neq 0$ if R is an integral domain, so $a_d b_e x^{d+e}$ is the highest-degree term of fg. □

(For a counterexample when R is not an integral domain, notice that $(2x+1)(3x+1) = 5x+1$ in $\mathbb{Z}/6\mathbb{Z}[x]$.)

Lemma 22

If R is an integral domain, so are $R[x]$ and $R[x_i]_{i \in I}$, and so is $R[[x_i]]$.

(The argument for power series is a bit different from the other ones – we have to look at the lowest power of x instead of the highest one.)

It turns out that polynomials also have a universal property:

Proposition 23

Suppose $f : I \to S$ is any function to a ring S and $\phi : R \to S$ is a morphism. Then there exists a unique ring morphism $\psi : R[x_i]_{i \in I} \to S$ such that $\psi|_R = \phi$ and $\psi(x_i) = f(i)$ for all $i \in I$.

In other words, what's important about polynomials is that we can substitute in any values in for our variables.

Lemma 24 (Division algorithm)

Let $f, g \in R[x]$, and suppose g is monic (or more generally that g's leading term is a unit in R). Then there exist unique polynomials $q, r \in R[x]$ such that $f = qg + r$ and $\deg(r) < \deg(g)$.

(This is proved by the usual long division algorithm, which works the same way as it does for dividing integers.)

Definition 25

A subset I of a ring R is an **ideal** (denoted $I \triangleleft R$) if it contains 0, is closed under addition, and for any $r \in R$ and $s \in I$ we have $rs \in I$.

Example 26

For any ring R, $\{0\}$ and R are both ideals (we call an ideal **proper** if $I \neq R$). Also, the set of even integers, which we can denote (2), is an ideal in $\mathbb{Z}$.

Example 27
If $\phi : R \to S$ is a morphism and $J \triangleleft S$, then the preimage $\phi^{-1}J = \{r \in R : \phi(r) \in J\}$ is an ideal in R. In particular, the **kernel** $\ker \phi = \phi^{-1}(0)$ of any morphism is an ideal.

Example 28
The image of an ideal is not necessarily an ideal. For example, take the embedding $\mathbb{Z} \hookrightarrow \mathbb{Q}$. Then (2) is an ideal in $\mathbb{Z}$, but it is not an ideal in $\mathbb{Q}$ because $\frac{1}{2} \times 2 = 1 \notin (2)$. (However, if $\phi : R \to S$ is **surjective** and $I \triangleleft R$, then $\phi I \triangleleft S$.)

Definition 29
Let $X \subseteq R$ be a subset. The **ideal generated by X**, denoted (X), is the set $(X) = \left\{\sum_{i=1}^{n} r_i x_i : x_i \in X, r_i \in R\right\}$. A **principal ideal** is an ideal generated by only one element.

We can check that this is indeed an ideal of R, and for any ideal $I \supset X$ we have $I \supset (X)$. And this explains the notation (2) from Example 26.

3 September 30, 2022

Last lecture, we introduced the concept of **ideals**, seeing that ideals behave well under inverse images (and also forward images if we have a surjective map). Today, we'll start with ways to construct new ideals from existing ones.

Definition 30
Let $I, J \triangleleft R$ be two ideals. The sum of the ideals is $I + J = \{r + s : r \in I, s \in J\}$, the intersection ideal $I \cap J$ is their set intersection, and the product ideal IJ is the set of finite sums $\{\sum_{i=1}^{n} r_i s_i : r_i, \in I, s_i \in J\}$.

We can see that $I + J$ is indeed an ideal by direct checking of the definitions – it's actually the smallest ideal containing $I \cup J$, because it contains both I and J and any such ideal must be closed under addition so it must include $I + J$. Making similar arguments for the others, we can check that $\boxed{IJ \subset I \cap J \subset I, J \subset I + J}$.

Example 31
In $R = \mathbb{Z}$, taking $I = (6)$ and $J = (10)$ yields $I + J = (2)$ (because $2 = 2 \cdot 6 - 10$, and any element in $I + J$ must be even), $I \cap J = (30)$, and $IJ = (60)$.

Remark 32. *Notice that* $2 = \gcd(6, 10)$ *and* $30 = \operatorname{lcm}(6, 10)$. *As a general heuristic, having* $I \supset J$ *can be thought of as having "*$I|J$*," $I + J$ as the "gcd" of the ideals, $I \cap J$ as the "lcm", and IJ as the "product." (In the case where we have principal ideals in* $\mathbb{Z}$*, we've just seen that this does make sense.)*

Definition 33
Two ideals I, J are **comaximal** if $I + J = R$.

(In the language of integers, this is being "coprime.") Because $R = (1)$, being comaximal is equivalent to being able to find $r \in I$ and $s \in J$ with $r + s = 1$.

Lemma 34

Any ideal $I \triangleleft \mathbb{Z}$ is of the form (n) for some n. Similarly, if K is a field and $I \triangleleft K[x]$, then $I = (f)$ for some polynomial $f \in K[x]$.

So in these special cases, the principal ideals are the only ones.

Proof. For the integer case, we either have $I = (0)$ or some positive integer in I (by closure under multiplication by -1). Let n be the minimal such integer in I; we claim that $I = (n)$. Indeed, by the division algorithm, we can write any $m \in I$ as $m = qn + r$ with $0 \leq r < n$, but this means that $r = m - qn \in I$ (by ideal closure properties) so it must be 0 by minimality of n, meaning that $n|m$.

The same proof works for $K[x]$ using the division algorithm for polynomials and choosing nonzero f of minimal degree – importantly, we can choose f to be monic because K is a field and we can multiply by the inverse of the leading coefficient. (So if $g = qf + r$ and $\deg r < \deg g$, we must have $r = 0$.) □

Example 35

Consider $I = (1407, 917) \triangleleft \mathbb{Z}$. Since $1407 = 917 + 490$, 490 must be in the ideal as well. Then $917 = 490 + 427$, so 427 is in the ideal too, and repeating this process yields that $63, 49, 14$, and 7 are in the ideal as well. But now 7 does generate the ideal, because working backwards we see that every other number we've considered is a multiple of 7 as well, and thus $I = (7)$.

This is the **Euclidean algorithm**, and we can notice that it also allows us to write 7 as a multiple of 1407 and 917 in a systematic way (by plugging in $14 = 63 - 49$, $49 = 427 - 6 \cdot 63$, and so on):

$$7 = 49 - 3 \cdot 14 = 4 \cdot 49 - 3 \cdot 63 = 4 \cdot 427 - 27 \cdot 63 = 31 \cdot 427 - 27 \cdot 490 = 31 \cdot 917 - 58 \cdot 490 = 89 \cdot 917 - 58 \cdot 1407.$$

(We'll see that such an expression can be useful for solving certain kinds of problems.)

Lemma 36

A ring R is the zero ring (0) if and only if it has exactly one ideal (namely R itself), since (0) and R are always ideals of R. Similarly, R is a field if and only if it has exactly two ideals, since this is equivalent to saying that any ideal containing (or generated by) a nonzero $r \in R$ also contains 1.

Definition 37

For any $I \triangleleft R$, the **quotient ring** is the set of cosets $R = \{r + I : r \in R\}$ with addition and multiplication defined in the usual way – $(r + I) + (s + I) = (r + s) + I$, and $(r + I)(s + I) = rs + I$ – and the additive and multiplicative identities $0 + I$ and $1 + I$, respectively.

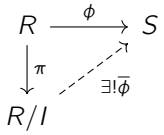
We can check that this is indeed well-defined, particularly for multiplication: if we chose different representatives r', s' instead of r, s (meaning that $r' - r, s' - s \in I$), we must check that $r's' - rs \in I$. But indeed, $r's' - rs = r'(s' - s) + s(r' - r)$ is the product of two terms in I and is thus in I, meaning that $rs + I = r's' + I$.

There is a (surjective) ring morphism $\pi : R \to R/I$, which we call the **quotient map**, sending any r to the coset $r + I$. And like with the other properties we've mentioned, the quotient ring has the following universal property:

Lemma 38

Let $I \triangleleft R$, and suppose $\phi : R \to S$ is a morphism such that $\phi(I) = \{0\}$. Then there is a unique ring morphism $\overline{\phi} : R/I \to S$ with $\phi = \overline{\phi} \circ \pi$.

The diagram that needs to commute is shown below – any ring with this property must be the quotient ring R/I. (The notation $\exists!$ means "exists a unique map.")



In particular, if we take I to be $\ker\phi$, we see that any surjective map $\phi : R \to S$ yields an **isomorphism** $\overline{\phi} : R/\ker\phi \to S$ (this is sometimes called the **first isomorphism theorem**).

Lemma 39

The ideals of $R \times S$ are exactly the ideals of the form $I \times J$, where $I \triangleleft R$ and $J \triangleleft S$, and $(R \times S)/(I \times J) \cong R/I \times S/J$ (through the map $(r, s) + I \times J \mapsto (r + I, s + J)$).

Proof. It's easy to see that any such $I \times J$ is indeed an ideal from the definition. Now suppose $K \triangleleft R \times S$ and that K contains some element (r, s). Then K also contains $(1, 0)(r, s) = (r, 0)$ and $(0, 1)(r, s) = (0, s)$, so we can verify that K is in fact of the form $(R \cap K) \times (S \cap K)$. □

Lemma 40

If $I \triangleleft R$, then $I[x] = \left\{\sum_{i=0}^{d} r_i x^i : r_i \in R\right\}$ is an ideal of $R[x]$. (But there are many other ideals in $R[x]$ in general.) We then have $R[x]/I[x] \cong (R/I)[x]$ (through the map $\sum r_i x^i + I[x] \mapsto \sum (r_i + I)x^i$).

Lemma 41

If $I \triangleleft R$, then there is a bijection between ideals of R/I and ideals of R containing I. In particular, if $\pi : R \to R/I$ is the quotient map, we can take an ideal $\overline{J}$ in R/I and get an ideal $\pi^{-1}\overline{J}$ of R containing I, and we can apply the surjective map π to get an ideal πJ of R/I from an ideal J of R containing I.

We just need to check that the composite of these two maps is the identity, which is left as an exercise to us. In particular, we then find that $(R/I)/(\pi J) \cong R/J$ through the map $r + J \mapsto (r + I) + \pi J$. The following type of argument can be useful in situations like this:

Corollary 42

For any $r, s \in R$, we have

$$R/(r, s) \cong (R/(r))/(s + (r)) \cong (R/(s))/(r + (s)),$$

so we can choose the order in which to mod out by elements in R.

We'll now turn to the Chinese remainder theorem:

Lemma 43
For any ideals $I, J \triangleleft R$, we have $R/(I \cap J) \cong (R/I) \times_{R/(I+J)} (R/J)$, where the map used in the relative product is $r + (I \cap J) \mapsto (r + I, r + J)$.

For example, applying this to Example 31 above yields $\mathbb{Z}/30\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z} \times_{\mathbb{Z}/2\mathbb{Z}} \mathbb{Z}/10\mathbb{Z}$.

Lemma 44
If I and J are comaximal, then $I \cap J = IJ$, so $R/IJ \cong R/(I \cap J) \cong R/I \times_R R/J \cong R/I \times R/J$.

For example, because 5 and 7 are coprime in $\mathbb{Z}$, we find that $\mathbb{Z}/35\mathbb{Z} \cong \mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/7\mathbb{Z}$. And the only additional thing here to prove is that anything in $I \cap J$ is also in IJ, and indeed for any $x \in I \cap J$ we can write (finding $r, s \in I, J$ so that $r + s = 1$) $x = xr + xs$, and both xr and xs are in IJ so x is also in the product ideal. Generalizing this to several ideals (and using induction) yields the result:

Proposition 45 (Chinese remainder theorem)
If $I_1, \cdots, I_n \triangleleft R$ are pairwise comaximal ideals (meaning that $I_i + I_j = R$ for any $i \neq j$), then $I_1 I_2 \cdots I_n = I_1 \cap \cdots \cap I_n$, and $R/I_1 \cdots I_n \cong R/(I_1 \cap \cdots \cap I_n) \cong R/I_1 \times \cdots \times R/I_n$.

4 October 3, 2022

Last week, we discussed rings and ideals, particularly looking at the Chinese remainder theorem. Today,

Definition 46
An ideal $I \triangleleft R$ is a **maximal ideal** if it is not the whole ring (it is **proper**) and not properly contained in any other proper ideal.

Notice that I is maximal if and only if R/I is a field (because the ideals of R/I are in correspondence with the ideals of R containing I, and we have a field if and only if there are exactly two ideals).

Definition 47
An ideal $I \triangleleft R$ is a **prime ideal** if it is proper, and whenever $rs \in I$, either $r \in I$ or $s \in I$. The set of prime ideals of R is denoted $\operatorname{Spec} R$.

We can see that I is prime if and only if R/I is an integral domain (because $rs \in I$ is the same as $rs + I$ being zero in R/I). Since any field is an integral domain, this means that any maximal ideal is prime.

Example 48
The prime ideals in $\mathbb{Z}$ are the zero ideal (0) and the ideals generated by prime numbers (p). (This explains the name "prime.") Indeed, every ideal is principal, and if we had an ideal (m) where $m = rs$, then we must either have $r \in S$ or $m \in S$, which can only occur if $r = m$ or $s = m$.

$\operatorname{Spec} R$ turns out to be a topological space rather than just a set, but we won't really go into that here. This set of prime ideals turns out to behave well under pullbacks:

Lemma 49

If $\phi : R \to S$ is a ring morphism, and $J \triangleleft S$ is a prime ideal, then $\phi^{-1}J$ is a prime ideal of R.

Proof. We already know that the pullback of an ideal is an ideal, so we just need to check that it is prime. We know that $\phi^{-1}J$ is proper, because $1 \in \phi^{-1}J$ would imply that $1 \in J$, which is not the case (because J is a prime ideal and thus proper). And if $rs \in \phi^{-1}J$, then $\phi(rs) \in J$, meaning that $\phi(r)\phi(s) \in J$, so either $\phi(r)$ or $\phi(s)$ is in J. So indeed either r or s is in $\phi^{-1}J$. □

This means we have a map $\phi^{-1} : \operatorname{Spec} S \to \operatorname{Spec} R$ for any ring morphism $\phi : R \to S$. However, the pullback of a maximal ideal is not necessarily maximal – the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is a ring morphism, but the preimage of (0) is (0), and (0) is maximal in $\mathbb{Q}$ but not in $\mathbb{Z}$. So prime ideals have better functoriality properties than maximal ones, and that's why the development of the subject has followed prime ideals.

Lemma 50

The prime ideals of a product ring $R \times S$ are of the form $I \times S$ and $R \times J$, where $I \triangleleft R$ and $J \triangleleft S$ are prime ideals.

(So the ideals in a product are just products of ideals, but the prime ideals are not just products of prime ideals.)

Proof. If $K \triangleleft R \times S$ is a prime ideal, then it must contain $0 = (0,1) \cdot (1,0)$, so it contains either $(0,1)$ or $(1,0)$. In the first case we must multiply by S and in the second we must multiply by R, and we can check that the other argument must be a prime ideal. □

In other words, we can rephrase this as a "disjoint union"

$$\operatorname{Spec} R \times S = \operatorname{Spec} R \sqcup \operatorname{Spec} S.$$

Remark 51. *However, we cannot extend this argument to an infinite product. Consider $\prod_{i=1}^{\infty} \mathbb{Q}$ – the kernel of the projection onto any factor is a prime ideal, but there are other prime ideals as well. If $\mathcal{X}$ is a collection of subsets of $\mathbb{Z}_{>0}$ (the index set here), define $I(\mathcal{X}) = \{(r_i) \in \prod_{i=1}^{\infty} \mathbb{Q} : \{i : r_i = 0\} \in \mathcal{X}\}$. In other words, given a subset of the positive integers, we get a subset of the ring. Then we can check what properties $\mathcal{X}$ must have to get a prime ideal (exercise, but for the projection π_i it would be all sets containing i), and we can check that such an $\mathcal{X}$ does exist.*

Lemma 52

Let I be an ideal of R. Then there is a bijection between prime ideals of R/I and prime ideals of R containing I.

Proof. We already know there is a bijection between the set of all ideals: if $\pi : R \to R/I$ is a quotient map sending J to πJ, with inverse map sending $\overline{J}$ to $\pi^{-1}\overline{J}$, we want to check that prime ideals are preserved. The latter has been shown in Lemma 49, so we just need to check that if J is prime, then πJ is also prime. But if $(r + I)(s + I) \in \pi J$, then $rs + I \in \pi J$ is equivalent to saying that $rs \in J$, meaning that either $r \in J$ or $s \in J$. Thus either $r + I \in \pi J$ or $s + I \in \pi J$, as desired. □

We'll now discuss the **noetherian property**, an important property of rings that holds for most of the rings that we will encounter.

Lemma 53

The following two properties of a ring R are equivalent:

1. Any ideal of R is finitely generated,
2. If $\mathcal{X}$ is any nonempty set of ideals of R, then $\mathcal{X}$ has a maximal element $I \in \mathcal{X}$ (meaning that it is not properly contained in any other ideal, but it's not required to contain every other ideal).

Proof. First assume (1) but assume (2) is false, so there is some set of ideals $\mathcal{X}$ with no maximal element. Take any $I_1 \in \mathcal{X}$; since property (2) is false, I_1 is not maximal and there is some $I_2 \in \mathcal{X}$ such that $I_1 \subsetneq I_2$. Similarly we can find $I_3 \in \mathcal{X}$ such that $I_2 \subsetneq I_3$; continuing on we get an infinite increasing chain of ideals in $\mathcal{X}$. Then the nested union $I = \bigcup_{i=1}^{\infty} I_i$ is also an ideal (indeed any element of I is in some I_n so multiplying by any other element keeps us in I_n and thus in I, and if we have two elements they are both in some sufficiently large I_n), so it is finitely generated by some set of elements $(r_1, \cdots, r_n)$. But each r_i is in some I_n, so there is some N such that $r_1, \cdots, r_n \in I_N$ and thus $I \subset I_N$ (we already have the whole union at step N). This contradicts the fact that $I_N \subsetneq I_{N+1}$, so (2) cannot be false.

On the other hand, suppose (2) is true. For any ideal $I \triangleleft R$, let $\mathcal{X}$ be the set of finitely generated ideals contained in I. This is a nonzero set of ideals because it contains (0), so it has a maximal element J_0. We know that $J_0 \subset I$, but if J_0 is not all of I then there is some $r \in I - J_0$, and then the ideal (J_0, r) is still finitely generated (so is in $\mathcal{X}$) and strictly contains J_0, which contradicts maximality. So we must actually have $J_0 = I$ and thus I itself is finitely generated. □

Definition 54

A ring R is **noetherian** if the properties in Lemma 53 hold.

(Noether and Abel are two mathematicians whose names are often not capitalized when used in math.)

Example 55

$\mathbb{Z}$ and $K[x]$ (for any field K) are noetherian (because any ideal is principal and thus generated by one element), and so is K for any field (because it only has two ideals generated by 0 and 1, respectively). However, the polynomial ring $\mathbb{C}[x_1, x_2, \cdots]$ is not noetherian because the ring generated by $(x_1, x_2, \cdots)$ is not finitely generated (any finite set of generators only involves finitely many of the x_is).

Remark 56. *Prompted by a question in class, the set of formal Laurent series $\mathbb{C}((x))$ is actually a field, because we've basically taken $\mathbb{C}[x]$ and added an inverse $\frac{1}{x}$ to the only element that isn't invertible.*

Lemma 57

If R and S are noetherian, then so is $R \times S$, and if R is noetherian and $I \triangleleft R$, then so is R/I.

(Both of these follow by our characterization of ideals of products and quotients.) As a warning, though, relative products $R \times_T S$ do not need to be noetherian even if R, S, and T are all noetherian.

Proposition 58 (Hilbert's basis theorem)

If R is noetherian, then so is $R[x]$ (and thus a polynomial ring over R in finitely many variables).

Proof. Let $I \triangleleft R[x]$ be an ideal. To "get back down to R," consider the set of elements

$$L_d = \{r \in R : r \text{ is the coefficient of } x^d \text{ of some } f \in I \text{ of degree at most } d\}.$$

(In other words, r is either zero or it's a leading coefficient.) This is an ideal of R, because we can add polynomials in I together and also multiply f by any $s \in R$ to get another degree d polynomial (or zero) now with x^d coefficient rs (or get zero, which is also in L_d). Furthermore, $L_d \subset L_{d+1}$ because we can always multiply any polynomial by x and preserve the "leading coefficient." By the noetherian property for R, we see that $L_N = L_{N+1} = \cdots$ for some sufficiently large N.

But by the noetherian property (in the other sense), there exist finitely many polynomials $f_{d,1}, \cdots, f_{d,s_d} \in I$ of degree at most d whose coefficients of degree d generate L_d. Specifically, for $d \geq N$, we can just take $s_d = s_N$ and $f_{d,i} = f_{N,i}x^{d-N}$ as a generating set for L_N. Now the ideal $J = (f_{i,j} : 0 \leq i \leq N, 1 \leq j \leq s_i)$ is a finitely generated ideal. Each $f_{i,j}$ is in I so $J \subset I$, but J is finitely generated and contains the $f_{i,j}$ for $i > N$ as well (since those are just the $f_{N,i}$s multiplied by a power of x). We now claim that $J = I$; otherwise, there is some $g \in I \setminus J$ of smallest degree d. Then removing its leading terms by subtracting off some term $\sum_j r_j f_{d,j}$ from J gets us another element in $I \setminus J$ with smaller degree, which is a contradiction. □

5 October 5, 2022

Last lecture, we mentioned two equivalent (and both useful) characterizations for being a noetherian ring, namely that (1) any ideal is finitely generated and (2) any non-empty collection of ideals has a maximal element. We saw that products and quotients of noetherian rings are still noetherian, and we also saw (by Hilbert's basis theorem) that polynomial rings (in finitely many variables) $R[x_1, \cdots, x_n]$ are noetherian if R is noetherian. It turns out that the formal power series ring $R[[x]]$ is also noetherian – the proof is similar as for polynomial rings, but we look at the lowest-degree terms rather than the highest-degree ones. Here's a reformulation of Hilbert's basis theorem:

Corollary 59

Suppose R is noetherian and $\phi : R \to S$ is a ring morphism. Then if S is finitely generated over R (meaning that there is a finite subset $X \subset S$, such that S has no proper subring containing $\text{im}(\phi)$ and X), then S is noetherian.

Proof. If we write $X = \{X_1, \cdots, X_n\}$, then there is a map $\psi : R[x_1, \cdots, x_n] \to S$ such that r maps to $\phi(r)$ and $X_1 \mapsto x_i$ (this is the universal property of the polynomial ring). This map should be surjective in order for X to generate S over R (otherwise the image would be a proper subring of S), and that means $S \cong R[x_1, \cdots, x_n]/\ker\phi$. But $R[x_1, \cdots, x_n]$ is noetherian, so the quotient is also noetherian. □

It will turn out that many rings that we will meet in "real life" are finitely generated over some simple ring, and thus being noetherian is a reasonable assumption.

Our next construction will be **rings of fractions**, and the example to keep in mind is to get from $\mathbb{Z}$ to $\mathbb{Q}$:

Definition 60

A subset $D \subset R$ is **multiplicative** if it contains 1 and is closed under multiplication.

It is easy to check that if $\phi : R \to S$ is a morphism and $D \subset R$ is multiplicative, then $\phi(D) \subset S$ is multiplicative. D will be the set of elements that can go into the denominator, and in general the fraction $\frac{r}{a}$ should be the same as

$\frac{s}{b}$ (so $br = as$). So we'll consider the set $R \times D$, and suppose

$$(r, a) \sim (s, b) \text{ if } c(br - as) = 0 \text{ for some } c \in D.$$

To make sure this definition makes sense, we need to check that we actually have an equivalence relation. Transitivity is the only hard part: if $(r, a) \sim (s, b) \sim (t, c)$, then

$$d(br-as) = 0, \quad e(cs-bt) = 0 \implies ecd(br-as)+eda(cs-bt) = 0 \implies ecdbr = edabt \implies ebd(cr-at) = 0.$$

And now we see why we need the factor of c in the definition – we wouldn't have transitivity otherwise. (The c is not actually necessary if we have an integral domain, though.) And now we can define $D^{-1}R$ to be the set of equivalence classes of $R \times D$ under $\sim$, and it's important to note that when we see $\frac{r}{a} = [(r, a)]$ we don't actually uniquely determine r and a.

Proposition 61

$D^{-1}R$ is a ring with $0 = \frac{0}{1}, 1 = \frac{1}{1}$, and addition and multiplication defined via

$$\frac{r}{a} + \frac{s}{b} = \frac{br + as}{ab}, \quad \frac{r}{a} \cdot \frac{s}{b} = \frac{rs}{ab}$$

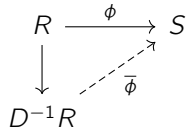
(which are elements of $D^{-1}R$ because $ab \in D$). Furthermore, we have a natural ring morphism $R \to D^{-1}R$ with $r \mapsto \frac{r}{1}$ (though it isn't always injective), and this map turns elements $d \in D$ into units with inverses $\frac{1}{d}$).

Proof. Most of the work here is in showing that all of these operations are well-defined. For example, if $\frac{r}{a} = \frac{r'}{a'}$ and $\frac{s}{b} = \frac{s'}{b'}$, we must check that $\frac{br+as}{ab} = \frac{b'r'+a's'}{a'b'}$; indeed, notice that $c(a'r - ar') = 0, d(b's - bs') = 0$ implies that $cd(a'b'br + a'b'as - abb'r' - aba's') = 0$ by pairing up the first and third terms, as well as the second and fourth terms. A similar check needs to be done for multiplication, and then we have to check ring axioms (such as associativity). Finally, checking that $R \to D^{-1}R$ is a ring morphism is more straighforward. □

We've mentioned that D is turned into units under the map $R \to D^{-1}R$, and it turns out this ring $D^{-1}R$ is the "cheapest way" to do so (this is a universal property):

Proposition 62

Suppose $D \subset R$ is multiplicative, and $\phi : R \to S$ is a morphism with $\phi(D) \subset S^\times$. Then there is a unique morphism $D^{-1}R \xrightarrow{\overline{\phi}} S$ such that the diagram below commutes.



We won't check the proof carefully, but to make sure the diagram commutes, we must define

$$\overline{\phi}\left(\frac{r}{a}\right) = \phi(r)\phi(a)^{-1}.$$

Verifying that this is well-defined and indeed a ring morphism is just working a bit with the definitions. And indeed if we have any other ring morphism ψ,

$$\psi\left(\frac{r}{a}\right)\phi(a) = \psi\left(\frac{r}{a}\right)\psi\left(\frac{a}{1}\right) = \psi\left(\frac{r}{1}\right) = \phi(r),$$

and because $a \in D$ we can invert $\phi(a)$ and we must have $\psi\left(\frac{r}{a}\right) = \phi(r)\phi(a)^{-1}$.

Example 63

If $0 \in D$, then $D^{-1}R$ will be the zero ring, because everything in $R \times D$ is equivalent. Indeed, we always have $\frac{r}{a} = \frac{0}{1}$ because $0 \cdot (r \cdot 1 - a \cdot 0) = 0$.

Example 64

If D is the set of elements of R that are not zero divisors, then D is multiplicative (since if ab were a zero divisor, then a would be one as well). In this case, $D^{-1}R$ is often called the **total quotient ring** QR. Then $R \hookrightarrow QR$ is injective, because $\frac{r}{1} = \frac{s}{1}$ if and only if $a(r-s) = 0$ for some $a \in D$, meaning $r = s$ because a is not a zero divisor. For example, $Q\mathbb{Z} = \mathbb{Q}$.

Lemma 65

If R is an integral domain (meaning $D = R \setminus \{0\}$), then QR is always a field.

Proof. For any fraction $\frac{r}{a} \neq \frac{0}{1}$ we have $r \neq 0$, so that fraction has inverse $\frac{a}{r}$. In particular, QR then ends up being the **smallest field containing** R. □

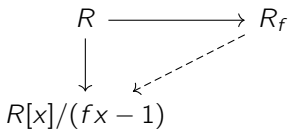
Example 66

For some explicit calculations, we have $Q(\mathbb{Z} \times \mathbb{Z}) = \mathbb{Q} \times \mathbb{Q}$ and $Q(\mathbb{C}[x]/(x^2)) = \mathbb{C}[x]/(x^2)$. In the latter case, we can check that $D = \{a + bx : a \neq 0\}$, which is the group of units of $\mathbb{C}[x]/(x^2)$ so we don't need to change the ring at all.

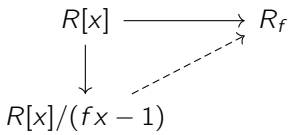
Example 67

For any $f \in R$, $\{1, f, f^2, \cdots\}$ is multiplicative, so we can define $R[1/f] = R_f = \{1, f, f^2, \cdots\}^{-1}R$.

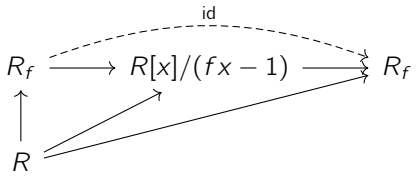
We claim that $R_f \cong R[x]/(fx-1)$. For one direction, we can construct the morphism $R_f \to R[x]/(fx-1)$ sending $\frac{r}{f^n}$ to rx^n (and we must check that this is indeed well-defined). But a more natural way to do this is to notice that we have a map $R \to R[x]/(fx-1)$ such that f^n is a unit for all n (because $f^n x^n = (fx)^n = 1$). So by the universal property (and factoring through the map $R \to R_f$) we also have a map $R_f \to R[x]/(fx-1)$.



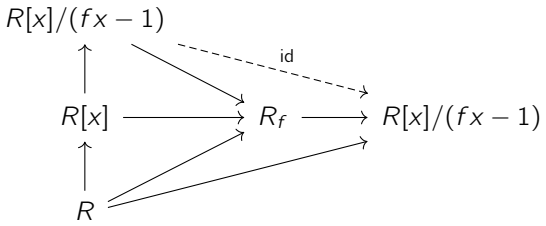
For the other direction, we can use the universal property of polynomial rings: the map $R \to R_f$ can be extended to a map $R[x] \to R_f$ by sending $x \mapsto \frac{1}{f}$, under which $fx - 1$ goes to zero, so we get a map from the quotient $R[x]/(fx-1) \to R_f$ by the universal property of quotients.



From here, we must show that these two maps are inverses, and we're basically doing that by "gluing commutative diagrams together." By our work above, the two triangles in the diagram below will commute. So the composition of the two maps $R_f \to R[x]/(fx-1) \to R_f$ is an extension of the map $R \to R_f$, but the identity map is the unique such map which does this, so our two morphisms do indeed compose to the identity.



For the other direction, we can construct the following diagram, where the maps in the middle row send x to $\frac{1}{f}$ to 0. Looking at the bottom half of the diagram, the universal property of $R[x]$ tells us that there is a unique map $R[x] \to R[x]/(fx-1)$, namely the quotient map. Thus that map must factor through the quotient $R[x]/(fx-1)$ (on the top half of the diagram), meaning the dashed map (the composition $R[x]/(fx-1) \to R_f \to R[x]/(fx-1)$) must again be the identity map.



Example 68

The ring $\mathbb{Z}[1/n]$ is the set of rational numbers whose denominator is a power of n, and $R[[x]][1/x] = R[[x]]_x$ is the Laurent series on R (containing elements $\sum_{n=N}^{\infty} a_n x^n$ where $a_n \in R$ and $N \in \mathbb{Z}$ is some finite "lowest power").

6 October 7, 2022

We constructed rings of fractions last time – for any multiplicative subset $D \subset R$ of a ring R (which contains 1 and is closed under multiplication), we can define $D^{-1}R$ to be the set of equivalence classes $[\frac{r}{d}] = [(r,d)]$ under the equivalence relation $(r,d) \sim (s,e)$ if $a(re-sd) = 0$ for some $a \in D$. In particular, when D is the set of all elements that are not zero divisors, we get the ring QR. (And the map $R \hookrightarrow QR$ is injective as long as D has no zero divisors.) After that, we defined $R_f = R[1/f]$ to be the ring of fractions defined by $D = \{1, f, f^2, \cdots\}$ – this is known as a **localization**.

We'll start today with some further constructions:

Example 69

If $D \subset R$ and $E \subset S$ are multiplicative subsets, then $D \times E \subset R \times S$ is also multiplicative (by checking the definition), and we get $(D \times E)^{-1}(R \times S) \cong D^{-1}R \times E^{-1}S$. In other words, the map $\frac{(r,s)}{(d,e)} \to \left(\frac{r}{d}, \frac{s}{e}\right)$ is well-defined and produces an isomorphism.

Example 70

Let $\phi : R \to S$ be a ring morphism, and let $D \subset R$ be multiplicative. Then $\phi D \subset S$ is multiplicative, so we can form $(\phi D)^{-1}S$, which is sometimes just abbreviated to $D^{-1}S$. Then we have the map $D^{-1}R \to D^{-1}S$ sending $\frac{r}{d}$ to $\frac{\phi(r)}{\phi(d)}$. So then by the universal property of rings of fractions, any map $R \to S \to D^{-1}S$ will factor through that map $D^{-1}R \to D^{-1}S$.

Next (this can be checked from the universal property for rings of fractions and polynomial rings), if we have S a polynomial ring,

$$(D^{-1}R)[x] \cong D^{-1}(R[x]).$$

(More directly, the isomorphism here sends $\sum_{i=0}^{n} \frac{r_i}{d_i} x^i$ to the fraction with a single denominator, which is a polynomial over $d_0 d_1 \cdots d_n$.) On the other hand, if we consider formal power series, we can consider $\mathbb{Z}[[x]][\frac{1}{2}]$ (in which we invert the elements $1, 2, 4, 8$, and so on), or we can consider $\mathbb{Z}[\frac{1}{2}][[x]]$. But the former contains only things of the form $\frac{\text{polynomial}}{2^n}$ but the latter contains elements like $1 + \frac{x}{2} + \frac{x^2}{4} + \frac{x^3}{8} + \cdots$. (In particular, we haven't described any universal property for the power series ring yet.)

Definition 71

Suppose $\mathfrak{p} \lhd R$ is a prime ideal. Then $R - \mathfrak{p}$ is a multiplicative set (if there were two things in $R - \mathfrak{p}$ whose product was not also in it, that would contradict the primeness of $\mathfrak{p}$). The **localization of R at $\mathfrak{p}$** is defined to be $R_{\mathfrak{p}} = (R - \mathfrak{p})^{-1}R$.

Example 72

For any prime ideal (p) in $\mathbb{Z}$ for a prime number p, $\mathbb{Z}_{(p)}$ inverts any integer not divisible by p, yielding the set of reduced rational numbers with no powers of p in the denominator. (This can be kind of confusing, because we defined $\mathbb{Z}_p = \mathbb{Z}[\frac{1}{p}]$ and that also looks like the p-adic numbers. So in general we should be careful about the difference between R_f and $R_{(f)}$, which invert completely different elements.) On the other hand, $\mathbb{Z}_{(0)} = \mathbb{Q}$.

Example 73

We have $\mathbb{C}[X]_{(0)} = \mathbb{C}(X)$, the field of rational functions of X. So then the localization $\mathbb{C}[X]_{(x)}$ is the set of (reduced) rational functions $\frac{p(X)}{q(X)} \in \mathbb{C}(X)$ for which $X \nmid q(X) \implies q(0) \neq 0$, or equivalently

$$\mathbb{C}[X]_{(x)} = \{f(x) \in C(x) : f \text{ has no pole at } 0\}.$$

This explains the term "localization" – this operation makes us care about the functions that make sense localized near zero.

Example 74

For any prime ideal $\mathfrak{p} \lhd S$, we know that $R \times \mathfrak{p} \lhd R \times S$ is a prime ideal, so we can define

$$(R \times S)_{R \times \mathfrak{p}} = (R \times S - R \times \mathfrak{p})^{-1}(R \times S) = (R \times (S - \mathfrak{p}))^{-1}(R \times S) = R^{-1}R \times (S - \mathfrak{p})^{-1}S,$$

and the first term in the product here is the zero ring so we just get back $S_{\mathfrak{p}}$ (and R doesn't do very much).

In particular, $R \times \mathfrak{p}$ often contains zero divisors, but localizing by it doesn't give us the zero ring – it just throws away some factors. (Containing zero is what immediately collapses the ring of fractions.)

Example 75

Let $D \subset R$ be multiplicative, and let $I \triangleleft R$ be an ideal. Then $D^{-1}I = \left\{\left[\frac{r}{a}\right] : r \in I, a \in D\right\} \triangleleft D^{-1}R$ is an ideal.

We say that I is **saturated** with respect to D if for any $r \in R$ and $a \in D$, having $ar \in I$ implies that we already had $r \in I$.

Lemma 76

If I is saturated with respect to D and $\frac{r}{a} \in D^{-1}I$, then $r \in I$.

(This is not obvious – $\frac{r}{a} \in D^{-1}I$ means that there's some representative equivalent to it with $r \in I$ and $a \in D$ – and saturation is required.)

Proof. Having $\frac{r}{a} \in D^{-1}I$ means that $\frac{r}{a} = \frac{s}{b}$ with $s \in I$ and $b \in D$. This means that $c(br - as) = 0$ for some $c \in D$. Since $cas \in I$, we must have $cbr \in I$ for some $c, b \in D$, so applying saturation to the element cb we see that $r \in I$. □

Proposition 77

Let $\phi : R \to D^{-1}R$ be the natural map.

1. For any ideal $J \triangleleft D^{-1}R$, the ideal $\phi^{-1}J \triangleleft R$ is saturated with respect to D, and $D^{-1}(\phi^{-1}J) = J$.
2. For any $I \triangleleft R$, $\phi^{-1}(D^{-1}I)$ is the smallest ideal containing I which is saturated with respect to D, and $D^{-1}(\phi^{-1}(D^{-1}I)) = D^{-1}I$.

Thus there is a bijection between ideals of $D^{-1}R$ and ideals of R that are saturated with respect to D, sending J to $\phi^{-1}J$ (and with inverse sending I to $D^{-1}I$). Additionally, if $\overline{D}$ is the image of D in R/I (which is multiplicative), there is an isomorphism $\overline{D}^{-1}(R/I) \cong D^{-1}R/D^{-1}I$, sending $\frac{r+I}{d+I}$ to $\frac{r}{d} + D^{-1}I$.

Proof sketch. To prove that $\phi^{-1}J$ is saturated, suppose $r \in R$, $a \in D$, and $\frac{ar}{1} \in J$. Then $\frac{1}{a} \cdot \frac{ar}{1} = \frac{r}{1} \in J$, meaning that $r \in \phi^{-1}J$. Meanwhile, $D^{-1}\phi^{-1}J$ is the set of $\frac{r}{a}$ such that $a \in D$ and $\frac{r}{1} \in J$. Any element $\frac{r}{a}$ of J is in that set because $\frac{r}{a} \in J$ implies $\frac{r}{1} \in J$, and for the other inclusion, $\frac{r}{a} = \frac{1}{a} \cdot \frac{r}{1}$.

(2) can be proved similarly and is left as an exercise to us. Finally, to show the bijection, our map $\overline{D}^{-1}(R/I)$ must send $\overline{D}$ to units and kill I, so we start with the map $R \to \frac{D^{-1}R}{D^{-1}I}$ sending r to $\frac{r}{1} + D^{-1}I$. This map sends I to zero, so by the universal property of the quotient our map factors through to a map $R/I \to \frac{D^{-1}R}{D^{-1}I}$. Now any element $d + I$ with $d \in D$ is sent to $\frac{d}{1} + D^{-1}I$, which has an inverse $\frac{1}{d} + D^{-1}I$, so D is indeed sent to units and thus by the universal property we get a map $\overline{D}^{-1}(R/I) \to \frac{D^{-1}R}{D^{-1}I}$, sending $\frac{r+I}{d+I}$ (for any $r \in R, d \in D$) to $\left(\frac{r}{1} + D^{-1}I\right)\left(\frac{d}{1} + D^{-1}I\right)^{-1} = \left(\frac{r}{1} + D^{-1}I\right)\left(\frac{1}{d} + D^{-1}I\right) = \frac{r}{d} + D^{-1}I$. So this map is indeed the one we wrote down in the statement.

To produce the map in the other direction, we first construct a map $D^{-1}R \to \overline{D}^{-1}(R/I)$, which we can get by first constructing the map $R \to \overline{D}^{-1}(R/I)$ sending r to $\frac{r+I}{1+I}$. Now something in D gets sent to a unit, so this map factors through to a map $D^{-1}R \to \overline{D}^{-1}(R/I)$. And then starting with an element $\frac{r}{d} \in D^{-1}I$ with $r \in I$, we are always sent to zero, so we get a map $D^{-1}R/D^{-1}I$ sending $\frac{r}{d} + D^{-1}I$ to $\left(\frac{r+I}{1+I}\right)\left(\frac{d+I}{1+I}\right)^{-1}$. Since the second term is in $\overline{D}$, we can simplify this to $\left(\frac{r+I}{1+I}\right)\left(\frac{1+I}{d+I}\right) = \frac{r+I}{d+I}$, showing that we do indeed have an inverse. □

Corollary 78

If R is noetherian, then any ring of fractions $D^{-1}R$ is noetherian, because any nonempty set of ideals in $D^{-1}R$ corresponds to a set of ideals in R and must have a maximal element.

Corollary 79

A prime ideal $\mathfrak{p} \triangleleft R$ is saturated with respect to D if and only if $\mathfrak{p} \cap D = \varnothing$. So there is a bijection between prime ideals of $D^{-1}R$ and prime ideals of R not intersecting D, given by the same maps $J \mapsto \phi^{-1}J$, $I \mapsto D^{-1}I$ for the forward and backwards directions. (We just need to check that we actually preserve primeness.) One interpretation of this is that $\operatorname{Spec}(D^{-1}R) \subset \operatorname{Spec}(R)$.

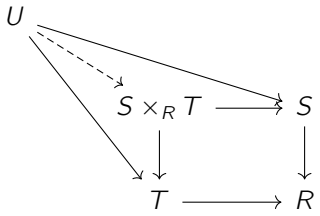
Finally, $\operatorname{Spec}(R_{\mathfrak{p}})$ is the set of prime ideals $I \in \operatorname{Spec} R$ such that $I \subset \mathfrak{p}$, so localization throws away prime ideals except those contained in $\mathfrak{p}$.

Remark 80. *There are some notes on* **completion** *that we should read to complete the problem set; they're not on the syllabus for the qualifying exam but they are for this class.*

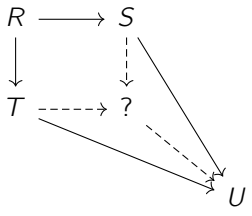
7 October 10, 2022

(As a logistical note, we should start submitting our homeworks on Gradescope now that there is a grader for the course helping out Lie.)

We'll discuss the last of our ring constructions today, the **tensor product**. The approach we'll take here is not the standard one – usually this is developed as a tensor product of modules, but there's some advantage in doing it for rings first. Recall that if we have two rings S and T with maps into R, then we formed the relative product $S \times_R T$ with the following universal property:



It makes sense to ask what happens if we reverse the arrows as shown below, forming a **pushout diagram**.



The answer is yes, and it's what gives rise to the tensor product:

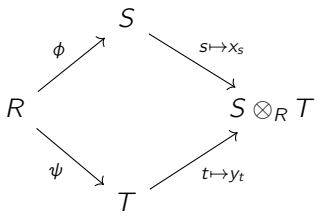
Definition 81

Suppose that we have rings R, S, T with maps $\phi : R \to S$ and $\psi : R \to T$. The **tensor product** $S \otimes_R T$ is

$$S \otimes_R T = R[x_s, y_t]_{s\in S, t\in T} / \left\langle x_{s_1+s_2} - x_{s_1} - x_{s_2}, x_{t_1+t_2} - x_{t_1} - x_{t_2}, x_{s_1 s_2} - x_{s_1} x_{s_2}, x_{t_1 t_2} - x_{t_1} x_{t_2}, x_{\psi(r)} - r, y_{\psi(r)} - r \right\rangle,$$

where these relations span over all $s_1, s_2 \in S$, $t_1, t_2 \in T$, and $r \in R$. (This definition does depend on ϕ and ψ but those are usually omitted in the notation.)

In other words, we form a polynomial ring in a huge number of variables (one variable for each element of s and t) and then mod out by the constraints required to make ϕ, ψ into ring morphisms. Then we can form the following commutative diagram, which commutes because R gets sent along as the identity map in both directions:



Unfortunately, this concrete definition is hard to work with directly, so we'll often appeal to other arguments. But first, we'll set up the notation that $s \otimes t = x_s y_t$ (these elements are called **pure tensors**) – notice that every element of $S \otimes_R T$ is a finite linear combination of pure tensors, because any polynomial is a finite sum of things of the form $r \prod x_{s_i}^{r_i} \prod y_{t_j}^{m_j}$. But it's usually not possible to write down a map from $S \otimes_R T$ by deciding where its pure tensors go, but it's hard to know whether we actually have $s_1 \otimes t_1 + s_2 \otimes t_2 + s_3 \otimes t_3 = 0$ (creating additional relations between the elements that are not immediately apparent).

Next, we find that

$$(s_1 + s_2) \otimes t = s_1 \otimes t + s_2 \otimes t, \quad s \otimes (t_1 + t_2) = s \otimes t_1 + s \otimes t_2$$

because of the first two relations multiplied by y_t and x_s respectively, and similarly

$$(s_1 \otimes t_1)(s_2 \otimes t_2) = s_1 s_2 \otimes t_1 t_2$$

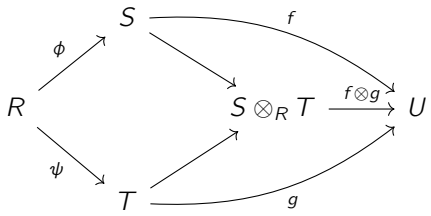
because of the next two relations, and then for any $r \in R$, we have

$$(\phi(r)s) \otimes t = s \otimes (\psi(r)t) = r(s \otimes t)$$

because both of these can be thought of as the monomial $r x_s y_t = x_s r y_t$.

Lemma 82 (Universal property of tensor product of rings)

Suppose we have a diagram as in the construction of $S \otimes_R T$ but with a different ring U, as shown below and with maps $f : S \to U$ and $g : T \to U$. Then there is a unique ring morphism $f \otimes g$ making the diagram below commute.

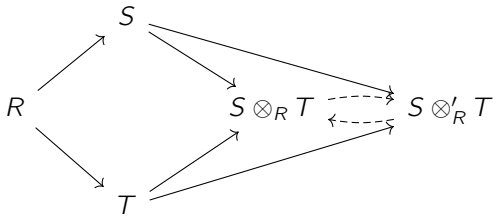


In general, we should think of using this universal property whenever we want to **map out of a tensor product**.

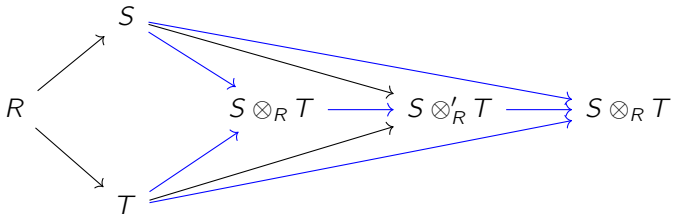
Proof. To make the two triangles on the right commute, we must have the map $R[x_s, y_t] \to U$ sending x_s to $f(s)$ and y_t to $g(t)$ for each $s \in S, t \in T$. We must then check that it factors through the quotient, showing that each generator of that ideal gets sent to zero. But that's true because f and g are linear, and this is the unique map because we've shown that every variable x_s and y_t must be sent to specific elements in U. □

The point now is that we only need this universal property and work with it instead of the original definition.

Remark 83. *By composing diagrams, we see that if $S \otimes'_R T$ also has the same property as $S \otimes_R T$, we get a canonical isomorphism. It looks as shown below:*



Indeed, the triangles marked in blue commute because each of the subtriangles inside them commute, and we have a unique map $S \otimes_R T \to S \otimes_R T$ which must be the identity map.

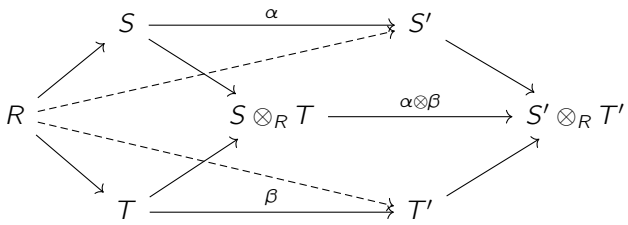


We'll now mention some other basic properties of tensor products:

Lemma 84

Suppose we have maps $\alpha : S \to S'$ and $\beta : T \to T'$ such that α commutes with maps $R \to S$ and $R \to S'$, and β commutes with maps $R \to T$ and $R \to T'$. Then there is a unique map $\alpha \otimes \beta : S \otimes_R T \to S' \otimes_R T'$ such that $s \otimes t$ maps to $\alpha(s) \otimes \beta(t)$.

Proof. By the universal property, we want to map from $S \otimes_R T$ out to $S' \otimes_R T'$, so we can just describe maps $S \to S' \otimes_R T'$ and $T \to S' \otimes_R T'$. One way to do that is to send s to $\alpha(s) \otimes 1$ and t to $1 \otimes \beta(t)$, but another way is to draw the diagram as follows:



Then the outer diagram commutes (the dashed maps come from the originally given maps $R \to S'$ and $R \to T'$), and then we can describe where any pure tensor goes via

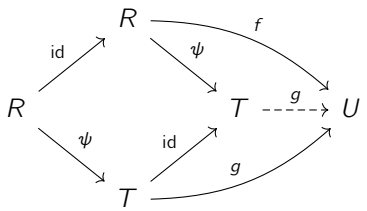
$$(\alpha \otimes \beta)(s \otimes t) = (\alpha \otimes \beta)(s \otimes 1 \cdot 1 \otimes t) = (\alpha \otimes \beta)(s \otimes 1)(\alpha \otimes \beta)(1 \otimes t) = (\alpha(s) \otimes 1)(1 \otimes \beta(t)) = \alpha(s) \otimes \beta(t)$$

□

Lemma 85

For any rings R, T and map $\psi : R \to T$, we have $R \otimes_R T \cong T$, with map sending $r \otimes t$ to $\psi(r)t$.

Proof. We can construct the maps explicitly between the two rings, but alternatively we can show that T has the universal property that $R \otimes_R T$ has. We must show that the following diagram commutes for any ring U, which is the same as showing that the outer diagram commutes. The dashed map exists and must be g if we want the bottom triangle to commute. But then the whole thing commutes because $f = g \circ \psi$ from the top triangle, so it's true for the big square as well.



□

Either one of those arguments works for the next result:

Lemma 86

We have an isomorphism $S \otimes_R T \cong T \otimes_R S$ sending $s \otimes t$ to $t \otimes s$.

Proof. We'll produce maps in both directions using the universal property of $S \otimes_R T$ and $T \otimes_R S$. Specifically, we get a unique map $\alpha : S \otimes_R T \to T \otimes_R S$ which sends $\alpha(s \otimes t) = \alpha(s \otimes 1 \cdot 1 \otimes t) = \alpha(s \otimes 1)\alpha(1 \otimes t) = (1 \otimes s) \cdot (t \otimes 1) = t \otimes s$. Similarly we get a unique map $\beta : (t \otimes s) = s \otimes t$. Then we can note that $\beta \circ \alpha(s \otimes t) = s \otimes t$, and being the identity on pure tensors also means we have the identity everywhere. The same works for $\alpha \circ \beta$. □

Definition 87

Let I be an ideal of S. Let $I \otimes_R T$ denote the ideal generated by pure tensors $s \otimes t$, where $s \in I$ (which contains all finite sums of pure tensors of this form).

We can indeed check that this is an ideal (exercise for us).

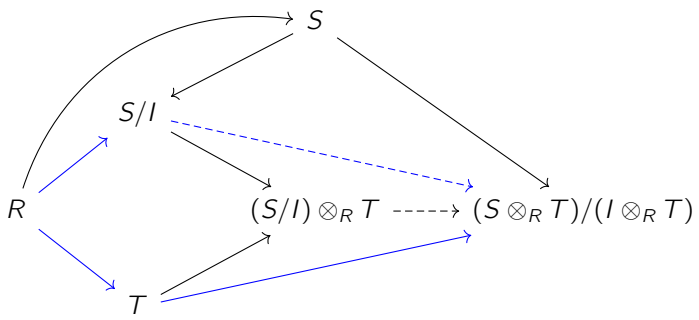
Lemma 88

For any $I \triangleleft S$, $I \otimes_R T$ is an ideal of $S \otimes_R T$, and we have

$$(S \otimes_R T)/(I \otimes_R T) \cong (S/I) \otimes_R T$$

by sending an element $s \otimes t + I \otimes_R T$ to $(s + I) \otimes t$.

Proof. We have a map $S \otimes_R T$ to $(S/I) \otimes_R T$ sending $s \otimes t$ to $(s+I) \otimes t$ (by applying Lemma 84 with the quotient map $S \to S/I$ and the identity map $T \to T$), and we can check that this factors through the quotient. For the other direction, we want to map out of the tensor product $(S/I) \otimes_R T$. We need to draw the following diagram:



The outside diagram commutes, and for any element $s \in I$, we get sent to 0 under the top right map. Thus (by universal property of the quotient) we factor to a unique (blue dashed) map $S/I \to (S \otimes_R T)/(I \otimes_R T)$ sending $s+I$ to $s \otimes 1 + I \otimes T$. Applying the universal property of the tensor product, we then get the dashed map $S/I \to S \otimes_R T/(I \otimes_R T)$ sending $s+I$ to $s \otimes 1 + I \otimes T$. We can thus construct the dashed map $S/I \otimes_R T$ to $(S \otimes_R T)/(I \otimes_R T)$. We see that

$$(s+I) \otimes t \mapsto (s \otimes 1 + I \otimes T)(1 \otimes t + I \otimes T) = (s \otimes t + I \otimes T).$$

Since we've kept track of what happens to pure tensors in both cases and we see that the maps are isomorphisms in both cases, we do indeed get the desired isomorphism. □

Combining the previous result with $S = R$ gives us the following result as well (which is easier than proving it from scratch):

Lemma 89
If $I \triangleleft R$ and $\psi : R \to T$, then $(R/I) \otimes_{R,\psi} T \cong T/\langle \psi(I) \rangle$

Lemma 90
If $D \subset S$ is multiplicative, then $D \otimes 1 = \{d \otimes 1 : d \in D\}$ is multiplicative, and

$$(D^{-1}S) \otimes_R T \cong (D \otimes 1)^{-1}(S \otimes_R T)$$

with map sending $\frac{s}{d} \otimes t$ to $\frac{s \otimes t}{d \otimes 1}$ and vice versa.

Combining this with another previous result gives us the following:

Lemma 91
For any multiplicative subset $D \subset R$ and any map $\psi : R \to 1$, we get $(D^{-1}R) \otimes_R T \cong D^{-1}T$, with map sending $\frac{r}{d} \otimes t$ to $\frac{\psi(r)t}{\psi(d)}$.

We also have a few other assorted useful facts:

Lemma 92
We have the polynomial ring isomorphism $S \otimes_R (T[x]) \cong (S \otimes_R T)[x]$.

Lemma 93

Tensor products are associative: we have an isomorphism $S \otimes_R (T \otimes_R U) \cong (S \otimes_R T) \otimes_R U$ sending $s \otimes (t \otimes u)$ to $(s \otimes t) \otimes u$.

Lemma 94

We have the isomorphism $S \otimes_R (T \times U) \cong (S \otimes_R T) \times (S \otimes_R U)$.

However, this last result may require modules to prove (there's no proof that Professor Taylor knows using only rings).

8 October 12, 2022

We'll spend today and next lecture on the final topic of rings, **factorization**. We know that integers can be written uniquely as powers of prime numbers and potentially a negative sign, and we now want to generalize this to other rings.

Definition 95

For any ring R and $r, s \in R$, we say that r **divides** s, denoted $r|s$, if $s = rt$ for some $t \in R$. An element $r \in R$ is **irreducible** if r is not a unit, but whenever $r = st$ in R, either s or t is a unit.

For example, the irreducibles in $\mathbb{Z}$ are primes and negative primes (because the only units are ± 1). And for any field K, the irreducibles in $K[x]$ are the irreducible polynomials in the usual sense.

Lemma 96

If R is a noetherian integral domain, then any nonzero $r \in R$ can be written as a product $r = u\pi_1 \cdots \pi_n$, where u is a unit and π_i are irreducibles.

Proof. This is a typical application of noetherianness: let X be the set of principal ideals (r) such that r does not have such a factorization. If X were empty, then we have the result that we want. Otherwise, X has some maximal element (r), and r is not irreducible because otherwise r would be its own factorization of the desired form. Thus $r = st$ for some s, t not units. But (s) contains (r), and if $(r) = (s)$ then we must have $s = ur \implies st = tur \implies 1 = tu$ (here we use that $st = r$ and that R is an integral domain), a contradiction because t is not a unit. So (s) cannot be in X by maximality of (r), and similar for (t), meaning that $s = u\pi_1 \cdots \pi_n$ and $t = v\pi_1' \cdots \pi_n'$. But this gives us a factorization $r = (uv)\pi_1 \cdots \pi_n \pi_1' \cdots \pi_n'$, a contradiction. □

Definition 97

Two elements $r, s \in R$ are **associates**, denoted $r \sim s$, if $r = su$ for some unit $u \in R^\times$.

(For example, p and $-p$ are associates in $\mathbb{Z}$, but different primes are not.) In the integers we know not only that every integer have a factorization, but also that it is unique up to order and switching signs (in other words, up to associates). More formally, if $r = u\pi_1 \cdots \pi_n = v\pi_1' \cdots \pi_m'$, where u and v are units and π_i, π_j' are all irreducible, then we know that $m = n$ and then for each i, we have $\pi_i \sim \pi_j'$ for some j. However, this is not true for all rings:

Example 98

Consider the ring $\mathbb{Z}[\sqrt{-5}]$. Then we have $6 = 2\cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$, but all of $2, 3, 1+\sqrt{-5}$, and $1-\sqrt{-5}$ are irreducible and 2 and $1 \pm \sqrt{-5}$ are not associates.

To check that 2 is not irreducible, suppose $2 = (a + b\sqrt{-5})(c + d\sqrt{-5})$ for some integers a, b, c, d. Taking the absolute value of these complex numbers (also called their norm), we see that $4 = (a^2 + 5b^2)(c^2 + 5d^2)$, which forces $b = d = 0$, and this leaves only trivial factorizations (with units) like $2 = (-1)\cdot(-2)$ and so on.

Definition 99

A ring R is a **unique factorization domain (UFD)** if it is an integral domain, every element r of $R \setminus \{0\}$ has a factorization $r = u\pi_1 \cdots \pi_n$ with u a unit and π_i irreducible, and such a factorization is unique, meaning that if $r = u\pi_1 \cdots \pi_n = v\pi'_1 \cdots \pi'_m$ for u, v units and π_i, π'_j irreducible, then $n = m$ and up to reordering we have $\pi_i \sim \pi'_i$.

(The existence of a factorization is usually not a problem – we just saw that it follows from noetherianness – so it is the uniqueness that matters.) For example, we know that $\mathbb{Z}$ is a UFD, and any field is a UFD because any nonzero element is already a unit (there are no irreducibles in a field). On the other hand, we've just shown that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

Lemma 100

Let R be an integral domain in which every element has a factorization. Then the following are equivalent:

1. The principal ideal (r) is prime if and only if r is irreducible.
2. If an element π is irreducible, then (π) is prime.
3. If π is irreducible and π divides rs, then π divides r or π divides s.
4. R is a UFD.

The point is that (2) or (3) are the weakest-looking conditions, so they are the easiest to check in practice.

Proof. Clearly (1) implies (2) (the latter is a weaker statement), and (2) and (3) are just restatements of each other.

To show (3) implies (4), suppose we have two factorizations $r = u\pi_1 \cdots \pi_n = v\pi'_1 \cdots \pi'_m$. We will prove that $n = m$ and $\pi_i \sim \pi'_i$ up to reordering by induction on m. For the case $m = 0$, each π_i would need to be a unit (which is not possible) so $n = 0$ and then the other statement is vacuous. Now if $m > 0$, π'_m divides $u\pi_1 \cdots \pi_n$, so by property (3) it either divides π_n or $u\pi_1 \cdots \pi_{n-1}$, meaning it either divides π_n, divides π_{n-1}, or divides $u\pi_1 \cdots \pi_{n-2}$. Repeating this process, we find that π_i is π'_m times a unit w for some i, so by reordering we can assume $i = n$. Substituting in $\pi_n = \pi'_m w$, we find that $(uw)\pi_1 \cdots \pi_{n-1} = v\pi'_1 \cdots \pi'_{m-1}$, so by induction $m-1 = n-1$ and up to order the remaining π_is and π'_is are associates.

To show that (4) implies (1), suppose R is a UFD. For one direction, the fact that if (r) is prime, then r is irreducible does not rely on the other properties (only on the fact that R is an integral domain). If r were not irreducible, then $r = st$ for non-units s, t, and by primeness of (r) we must have $s \in (r)$ or $t \in (r)$. But then we get the same argument as before: without loss of generality we have $s = ru$, so $st = rut \implies 1 = ut$, which means t is a unit. For the other direction, suppose r is irreducible. Then if $st \in (r)$ for some s, t, then $st = rx$ for some $x \in R$. Writing out the factorizations $s = u\pi_1 \cdots \pi_n$, $t = v\pi'_1 \cdots \pi'_m$, $x = w\pi''_1 \cdots \pi''_p$ and remembering that r is irreducible, we then get

$$s = u\pi_1 \cdots \pi_n v\pi'_1 \cdots \pi'_m = wr\pi''_1 \cdots \pi''_p.$$

But by unique factorization of st, we see that $r \sim \pi_i$ or π_i' for some i, meaning that r divides s or r divides t. Thus either $s \in (r)$ or $t \in (r)$, which proves that (r) is prime. □

Definition 101

A ring R is a **principal ideal domain (PID)** if it is an integral domain and any ideal $I \triangleleft R$ is principal (meaning that $I = (r)$ for some $r \in R$).

For example, the division algorithm tells us that $\mathbb{Z}$ and $K[x]$ are principal ideal domains.

Lemma 102

Suppose R is noetherian. Then R is a PID if and only if R is a UFD in which each nonzero prime ideal is a maximal ideal.

In particular, this tells us that $K[x]$ is always a UFD.

Proof. The reverse direction is on our homework as an exercise. For the forward direction, suppose R is a PID and $\pi \in R$ is irreducible. We wish to show that (π) is a prime ideal so that we can apply Lemma 100. If $rs \in (\pi)$, then $(r, \pi) = (t)$ for some t because we have a PID, and in particular $\pi = tu$ (because t divides π). By irreducibility, there are two cases: if u is a unit, then the ideals (t) and (π) are the same, so $r \in (t) \implies r \in (\pi)$, which is what we want. On the other hand, if t is a unit, then r and π must generate the whole ring $R<$ so we have $1 = ar + b\pi$ for some $a, b \in R$. Multiplying by s, we find that $s = a(rs) + (bs)\pi$, and now both terms on the right-hand side are divisible by π and thus π divides s, meaning $s \in (\pi)$.

It now remains to check that any nonzero prime ideal is maximal. Suppose we have two prime ideals $\mathfrak{p} \subset \mathfrak{q}$ with $\mathfrak{p} \neq 0$. If we have a PID, then $\mathfrak{p} = (\pi)$ and $\mathfrak{q} = (\pi')$, so $\pi = \pi' u$ for some unit u. But because π is irreducible and π' isn't a unit, u must be a unit and that means $(\pi) = (\pi')$. So we cannot have one prime ideal strictly contain another. Thus any nonzero prime ideal cannot be strictly contained in a maximal ideal (which would be prime), and thus the ideal itself must be maximal. □

Lemma 103

Let R be a UFD and let $r, s \in R$. The **greatest common divisor** $\gcd(r, s) \in R$ is the element such that $\gcd(r, s) | r$, $\gcd(r, s) | s$, and it is "greatest" in the sense that whenever $t | r$ and $t | s$, we also have $t | \gcd(r, s)$. Such a gcd always exists and is unique up to associates.

We can also similarly define gcds of more than two elements in an analogous way, as well as lcms.

Proof. Suppose $r = u\pi_1 \cdots \pi_n$ and $s = v\pi_1' \cdots \pi_m'$. Reorder so that $\pi_1 \sim \pi_1'\pi_2 \sim \pi_2', \cdots, \pi_p \sim \pi_p'$, but for any $i, j > p$ we have $\pi_i \not\sim \pi_j'$. Then we can set $\gcd(r, s) = \pi_1 \cdots \pi_p$ and check that all of the properties hold. □

Definition 104

We say that two elements $r, s \in R$ in a UFD are **coprime** if $\gcd(r, s) = 1$.

Definition 105

Let R be a UFD. A polynomial $f \in R[x]$ is **primitive** if the gcd of all of its coefficients are 1.

It will turn out that the product of two primitive polynomials is still primitive, and we'll see that and some related ideas next time.

9 October 14, 2022

Last lecture, we introduced unique factorization domains, which are integral domains where every element can be written as a unit times a bunch of irreducibles unique up to ordering and associates. We showed that being a UFD is equivalent to being an integral domain where factorizations exist and (π) is a prime ideal for every irreducible π. In a UFD, it makes sense to talk about the gcd and lcm of two elements, saying whether they're coprime, and so on. In the polynomial ring case, we said that a polynomial f is **primitive** if the coefficients have no common factor. This will play a role in factorization in the following way:

Lemma 106 (Gauss)

Let R be a UFD. Then if $f, g \in R[x]$ are primitive, then fg is primitive as well.

Proof. Suppose otherwise, so that there exists some irreducible $\pi \in R$ such that π divides all coefficients of fg. Since (π) is prime, we can define $\overline{f}$ and $\overline{g}$ to be the images of f and g in $R/(\pi)[x]$, which is the polynomial ring over an integral domain. But $\overline{f}\overline{g} = 0$ in $R/(\pi)[x]$ implies that either $\overline{f} = 0$ or $\overline{g} = 0$, which would contradict f and g being primitive. □

Lemma 107

Let R be a UFD. Then a polynomial $f \in R[x]$ is irreducible if and only if the following condition holds: either $f \in R$ and f is irreducible in R, or $f \in R[x]$ is primitive with f irreducible over the **fraction** field $Q(R)[x]$.

Proof. For the forward direction, suppose f is irreducible. Then if $f \in R$ and $f = gh$ in R, then either g or h is a unit in $R[x]^\times$, but all units in $R[x]^\times$ are elements of R. This means f is indeed irreducible in R. On the other hand, if $f \in R[x] \setminus R$ and some irreducible $\pi \in R$ divides all of the coefficients of f, then we can write $f = \pi \cdot (f/\pi)$, and neither π nor f/π can be a unit in $R[x]^\times = R^\times$. Thus f is indeed primitive in this case, and now we must check that it is irreducible over $Q(R)[x]$. If we can write $f = gh \in Q(R)[x]$ with degrees of g and h positive, we can pull out factors in the denominator of g and h by writing $g = a\tilde{g}$ with $a \in Q(R)$ and $\tilde{g} \in R[x]$ primitive (basically multiply by the lcm of the denominators), and similarly $h = b\tilde{h}$. Thus

$$f = (a\tilde{g})(b\tilde{h}) = ab\tilde{g}\tilde{h},$$

and by Gauss's lemma $\tilde{g}\tilde{h} \in R[x]$ is primitive. But now ab cannot have a denominator (we can always cancel common factors in a fraction in a UFD) – indeed, if there were some irreducible π in the denominator of ab, in order for $ab\tilde{g}\tilde{h} = f$ to be an element of $R[x]$ we must have π divide $\tilde{g}\tilde{h}$, which contradicts $\tilde{g}\tilde{h}$ being primitive. Thus $ab \in R$, and $f = (ab\tilde{g})\tilde{h}$ would give us a factorization in $R[x]$. This contradicts f being irreducible, so we must actually have irreducibility over $Q(R)[x]$ as desired.

The other direction is more straightforward: if $f \in R$ is irreducible and $f = gh$ in $R[x]$, then g and h must be in R, meaning g or h is in $R^\times$, so f is also irreducible in $R[x]$. And if f is primitive and irreducible in $Q(R)[x]$, then for any factorization $f = gh$ in $R[x]$, then thinking about all of those terms in $Q(R)[x]$ we can choose g to be an element of $Q(R) \cap R[x] = R$ without loss of generality. But because f is primitive, this can only occur if g is a unit in $R^\times = R[x]^\times$, and thus f is again irreducible. □

Theorem 108

If R is a unique factorization domain, so is $R[x]$ (and thus $R[x_1, \cdots, x_n]$ for any n).

Proof. To prove existence of a factorization, first notice that $Q(R)[x]$ is a UFD because $Q(R)$ is a field. So suppose we have $f = u\pi_1 \cdots \pi_n$ with $u \in Q(R)[x]^\times$, where $\pi_i \in Q(R)[x]$ are irreducible. But without loss of generality (by putting terms into u) we can say that π_i is primitive, so each $\pi_i \in R[x]$ is irreducible by our previous result. But then by the same logic as before, if $\pi_1, \cdots, \pi_n$ are primitive, so is their product. Thus u must be in R. This gives us a factorization of f in $R[x]$ as well.

For uniqueness, we'll instead check the fact that if $\pi \in R[x]$ is irreducible, then (π) is prime. Again by our previous result, we can check two cases: one case is where π is an irreducible element of R. Then $R[x]/(\pi) = R/(\pi)[x]$, and because $R/(\pi)$ is an integral domain so is $R/(\pi)[x]$. Thus $(\pi) \triangleleft R[x]$ is prime. The other case is where $\pi \in R[x] \setminus R$ is primitive and irreducible in $Q(R)[x]$. Suppose $\pi | fg$. Then since $Q(R)[x]$ is a UFD, we know that π divides either f or g in $Q(R)[x]$; without loss of generality let it be f. This means $f = \pi h$ for some $h \in Q(R)[x]$, and we can write $h = a\tilde{h}$ for some $a \in Q(R)$ and some primitive $\tilde{h} \in R[x]$. Then $f = a(\pi\tilde{h})$, but $\pi\tilde{h}$ is primitive so a cannot have any factors in the denominator. This means $a \in R$ and h was already in $R[x]$, meaning $\pi | f$ in $R[x]$ as well. That proves again that (π) is prime. □

We'll finish by discussing some other tricks for showing irreducibility of polynomials:

- The polynomial $x^3 - 3x + 1$ is irreducible in $\mathbb{Q}[x]$. Indeed, because it is primitive, irreducibility in $\mathbb{Q}[x]$ is equivalent to irreducibility in $\mathbb{Z}[x]$, and if it had a factorization there must be a linear factor, meaning that $x^3 - 3x + 1 = (x^2 + ax + b)(x + c)$ for integers a, b, c. But then $bc = 1$ implies that we'd either have $c = 1$ or $c = -1$, and ± 1 are both not roots of $x^3 - 3x + 1$. More generally, we get restrictions on the coefficients given primarily by the constant term if we're trying to look at irreducibility in $\mathbb{Z}[x]$.
- Next, $x^3 + x + 105$ is irreducible in $\mathbb{Q}[x]$. We can make a similar argument as before, but there's another trick here: if $x^3 + x + 105$ were reducible in $\mathbb{Z}[x]$, it would also be reducible in $\mathbb{Z}/(p)[x]$ (by reducing each polynomial individually). But reducing mod 2 gives us $x^3 + x + 1$, and now by the argument from above we see that this has no roots and thus cannot have a linear factor.

  However, we should be a bit careful with the leading coefficient here. For example, $2x^2 + x = x(2x + 1)$ is reducible, but mod 2 this is the polynomial x, which is irreducible. (The problem is that what was originally a valid irreducible factor becomes a unit.) So we just need to make sure we don't decrease the degree when we make such a reduction.

We'll now mention a less straightforward check for irreducibility:

Lemma 109 (Eisenstein's criterion)

Suppose R is an integral domain, $\wp \triangleleft R$ is a prime ideal. Suppose $f(x) = f_0 + f_1 x + \cdots + f_d x^d \in R[x]$. Suppose $f_d \notin \wp$, $f_i \in \wp$ for all $i < d$, and $f_0 \notin \wp^2$. Then f is not a product of two lower-degree polynomials.

To explain the technicality of the last point, notice that $2x+6$ is not irreducible, but it does satisfy the criterion for $\wp = (3)$ in $R = \mathbb{Z}$. For example, this shows that $x^4 + 10x + 5$ is irreducible by applying Eisenstein's criterion with the ideal (5).

Proof. Suppose $f = gh$ in $R[x]$, with $g(x) = g_0 + g_1x + \cdots + g_ex^e$ and $h(x) = h_0 + h_1x + \cdots + h_{d-e}x^{d-e}$ and $\deg g, \deg h < \deg f$. Comparing leading terms, we know that $g_e h_{d-e} = f_d \notin \wp$, so $g_e \notin \wp$ and $h_{d-e} \notin \wp$ (because $\wp$ is an ideal). Choose i, j minimal so that $g_i, h_j \notin \wp$. Then the coefficient f_{i+j} will contain a term g_ih_j, which is not in $\wp$ (this is the only place where we use that $\wp$ is a **prime** ideal), but it will have other terms $g_{i-k}h_{j+k}$ and $g_{i+k}h_{j-k}$, which will both be in $\wp$ because g_{i-k} and h_{j-k} are in $\wp$ by minimality of i, j. So we have a sum of terms in $\wp$ plus something not in $\wp$, which cannot be in $\wp$. This is only possible if $i + j = d$, and thus $i =$ and $j = d - e$. This would imply that all coefficients except the leading one are in $\wp$. But then $f_0 = g_0h_0 \in \wp^2$, which contradicts our original assumption. Thus f cannot be written as this product. □

Corollary 110

The p-cyclotomic polynomial $\phi_p(x) = \frac{x^p-1}{x-1} = x^{p-1} + x^{p-2} + \cdots + 1$, is irreducible in $\mathbb{Z}[x]$, because it's irreducible if and only if $\phi_p(x+1)$ is irreducible. And $\phi_p(x+1) = \frac{(x+1)^p-1}{x}$ has all coefficients divisible by p except the leading one and with constant term p, so it satisfies Eisenstein's criterion with the ideal (p).

10 October 17, 2022

We've finished our discussion on rings now, and we'll take a week of lectures to talk over the basics of category theory. A good reference for us would be Mac Lane's "Categories for the Working Mathematician."

Fact 111

We'll assume the existence of a **universe** $\mathcal{U}$, which is a set with the following properties:

- If $X \in Y$ and Y is in $\mathcal{U}$, then X is in $\mathcal{U}$.
- If $X, Y \in \mathcal{U}$, then $\{X, Y\} \in \mathcal{U}$.
- If $X \in \mathcal{U}$, then the powerset $\mathcal{P}(X)$ is in $\mathcal{U}$.
- The infinite set $\{0, 1, 2, 3, \cdots\}$ is in $\mathcal{U}$.
- If $X \in \mathcal{U}$ and $f : X \to \mathcal{U}$ is some function, then $\mathrm{Im}(f) \in \mathcal{U}$

(The point is that the axioms of a universe give us models for the axioms of set theory: elements of $\mathcal{U}$ give us a model of ZFC. But ZFC itself doesn't give us the existence of such a $\mathcal{U}$ – we have to add some additional axioms.) We'll call any $x \in \mathcal{U}$ a **small set**, but we shouldn't worry too much about this detail.

Definition 112

A **category** $\mathcal{C}$ consists of two sets $\text{ob}(\mathcal{C}) \subset \mathcal{U}$ (called the **objects**) and $\text{mor}(\mathcal{C}) \subset \mathcal{U}$ (called the **morphisms**), together with functions $\text{dom} : \text{mor}(\mathcal{C}) \to \text{ob}(\mathcal{C})$ and $\text{cod} : \text{mor}(\mathcal{C}) \to \text{ob}(\mathcal{C})$ (that is, for every morphism we can specify its domain and codomain), $\text{Id} : \text{ob}(\mathcal{C}) \to \text{mor}(\mathcal{C})$ (identity maps on each object), and a way to compose morphisms $\circ : \{(f, g) \in \text{mor}(\mathcal{C}) \times \text{mor}(\mathcal{C}) : \text{cod}(g) = \text{dom}(f)\} \to \text{mor}(\mathcal{C})$, such that the following axioms hold:

- $\text{dom}(\text{Id}_X) = X = \text{cod}(\text{Id}_X)$ for any X,
- $f \circ \text{Id}_X = f = \text{Id}_Y \circ f$ if f has domain X and codomain Y. (The set of morphisms starting at X and ending at Y will be denoted $\text{Hom}_{\mathcal{C}}(X, Y)$, and such morphisms may be written $f : X \to Y$ or $X \xrightarrow{f} Y$.)
- Associativity holds: $f \circ (g \circ h) = (f \circ g) \circ h$ if $\text{cod}(h) = \text{dom}(g)$ and $\text{cod}(g) = \text{dom}(f)$.

Remark 113. *Here, we don't want to think of "containment" or "elements of objects," which is why we require codomains and domains to be equal during composition. The point is to think as abstractly as possible.*

Definition 114

A morphism $f : X \to Y$ is an **isomorphism** if there exists a morphism $g : Y \to X$ with $f \circ g = \text{Id}_Y$ and $g \circ f = \text{Id}_X$. If such a morphism exists, it is unique (because if g, g' have this property then $g = g \circ f \circ g' = g'$), and we call this map the **inverse** of f and denote it f^{-1}.

Definition 115

A morphism $f : X \to Y$ is an **epimorphism** (sometimes **epi**) if whenever we have maps $g, h : Y \to Z$ with $g \circ f = h \circ f$ (and that composition makes sense), then $g = h$. $f : X \to Y$ is a **monomorphism** (sometimes **mono**) if whenever $g, h : Z \to X$ with $f \circ g = f \circ h$, then $g = h$.

This last definition makes more sense with the following concrete example:

Example 116

We have a category of sets (denoted **Set**), in which the objects $\text{ob}(\mathcal{C})$ are the small sets (so everything in $\mathcal{U}$) and the morphisms $\text{mor}(\mathcal{C})$ are the set of functions from one small set to another. Then domain, codomain, identity, and composition mean what they usually do for functions. In this category, we can check that being an epimorphism is equivalent to being **surjective**, and being a monomorphism is equivalent to being **injective**.

Example 117

The category **Grp** of small groups (groups whose elements form a small set) and group homomorphisms is defined with the usual composition of homomorphisms and so on, and being an epimorphism (resp. monomorphism) again corresponds to surjectivity and injectivity. The same is true for the category **Ring** of small rings and ring homomorphisms.

Example 118

Small topological spaces and continuous functions also form a category **Top**. Then monomorphisms are still injective, but epimorphisms don't need to be surjective. For example, the map $\mathbb{Q} \hookrightarrow \mathbb{R}$ (with the usual topology on $\mathbb{R}$ and either the subspace or discrete topology on $\mathbb{Q}$) is an epimorphism but not surjective.

Example 119

We will denote by **K-vect** the category of small K-vector spaces and K-linear maps and **Ab** the category of abelian groups and group homomorphisms – these are also indeed categories.

Definition 120

A **covariant functor** $F : \mathcal{C} \to \mathcal{D}$ between categories consists of the two functions $F : \text{ob}(\mathcal{C}) \to \text{ob}(D)$ and $F : \text{mor}(\mathcal{C}) \to \text{mor}(D)$, such that $F(\text{dom}(f)) = \text{dom}(F(f))$, $F(\text{cod}(f)) = \text{cod}(F(f))$, $F(\text{Id}_X) = \text{Id}_{F(X)}$, and $F(f \circ g) = F(f) \circ F(g)$. A **contravariant functor** is very similar but instead satisfies $F(\text{dom}(f)) = \text{cod}(F(f))$, $F(\text{cod}(f)) = \text{dom}(F(f))$, $F(\text{Id}_X) = \text{Id}_{F(X)}$, and $F(f \circ g) = F(g) \circ F(f)$.

We can think of covariant functors as **preserving the direction of arrows** and contravariant functors as **reversing the direction of arrows**: indeed, for a covariant functor we have $X \xrightarrow{f} Y$ becoming $F(X) \xrightarrow{F(f)} F(Y)$, but for a contravariant one we have $X \xrightarrow{f} Y$ becoming $F(X) \xleftarrow{F(f)} F(Y)$.

Example 121

We have a **forgetful functor** from **Grp** to **Set**, in which we forget about the group structure and think about the groups and homomorphisms as just sets and functions. We similarly have a forgetful functor **Ab** to **Grp**, as well as one from **K-vect** to **Ab**.

Example 122

The map H^i : **Top** $\to$ **Ab** sending X to $H^i(X, \mathbb{Z})$ is a contravariant functor (and similarly H_i is a covariant functor). Fundamental groups are a bit more tricky – the map π_1 sends **pointed** topological spaces to groups, but **Top** itself does not contain information about base points. And there is no category of "group up to isomorphism" because we don't know how to map between those objects.

Example 123

GL_n : **Ring** $\to$ **Grp** is a functor that sends any ring R to $\text{GL}_n(R)$.

Example 124

For any category $\mathcal{C}$, we have the identity functor $\text{Id}_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$.

Example 125

We have a contravariant functor $* :$ **K-vect** $\to$ **K-vect** sending V to its dual space V^*.

Definition 126

A functor $F : \mathcal{C} \to \mathcal{D}$ is **faithful** if it keeps distinct morphisms distinct – in other words, the map $\text{Hom}_{\mathcal{C}}(X, Y)$ to either $\text{Hom}_{\mathcal{C}}(F(X), F(Y))$ (if covariant) or $\text{Hom}_{\mathcal{C}}(F(Y), F(X))$ (if contravariant) is injective. Similarly, a functor is **full** if that map is surjective.

For example, forgetful functors are faithful but not full, the identity functor is fully faithful, and the duality functor $*$ is faithful but not full. Indeed, if the duals f^* and g^* of two linear maps $f, g : V \to W$ are equal, then $\lambda \circ f = \lambda \circ g$ for all functionals $\lambda : W \to K$, but if $f(v) \neq g(v)$ then there would be some λ such that $\lambda(f(v)) \neq \lambda(g(v))$, so $\lambda \circ f$ would not be equal to $\lambda \circ g$. But there are infinite-dimensional vector spaces where the duals are much bigger than the original spaces, so $*$ is not full. On the other hand, $*$ is fully faithful in the category of finite-dimensional vector spaces **Fin-K-vect** (because the double dual is always equal to itself).

Definition 127

We call $\mathcal{C} \subset \mathcal{D}$ a **subcategory** if $\text{ob}(\mathcal{C}) \subset \text{ob}(\mathcal{D})$ and $\text{mor}(\mathcal{C}) \subset \text{mor}(\mathcal{D})$, where $\mathcal{C}$ preserves domain, codomain, identity, and composition from $\mathcal{D}$.

For any subcategory $\mathcal{C} \subset \mathcal{D}$, we get an **inclusion functor** $I : \mathcal{C} \to \mathcal{D}$ (for example we have one from **Fin-K-vect** to **K-vect**), which is always faithful. If I is also full, we call $\mathcal{C}$ a **full subcategory** (for example **Fin-K-vect** is a full subcategory of **K-vect**).

We can also compose functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{E}$ to get a functor $G \circ F : \mathcal{C} \to \mathcal{E}$. Everything up to this point should be familiar in formalism with what we've already done, but now we have something a layer deeper:

Definition 128

Let $F, G : \mathcal{C} \to \mathcal{D}$ be two functors. A **natural transformation** $\phi : F \to G$ is a map which, for each object $X \in \mathcal{C}$, yields a morphism $\phi_X : F(X) \to G(X)$ which is compatible with all morphisms in $\mathcal{C}$, meaning that the following diagram commutes:

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

Example 129

There is a natural transformation det : GL_n to GL_1. In other words, for any ring R, we have a map sending $\text{GL}_n(R)$ to $\text{GL}_1(R)$ (the set of units of R) by taking the determinant.

That determinant is defined as a universal polynomial in the entries, and det is a natural transformation. Indeed, for any map $f : R \to S$ we get a map $\text{GL}_n(R) \to \text{GL}_n(S)$ (applying f to each entry) such that "taking the determinant is the same before or after taking the morphism:"

$$\begin{array}{ccc} \text{GL}_n(R) & \xrightarrow{\det} & \text{GL}_1(R) \\ \downarrow{\scriptstyle \text{GL}_n(f)} & & \downarrow{\scriptstyle \text{GL}_1(f)} \\ \text{GL}_n(S) & \xrightarrow{\det} & \text{GL}_1(S) \end{array}$$

We'll see some more examples of this next time!

11 October 19, 2022

Last lecture, we introduced categories $\mathcal{C}$ and functors $F : \mathcal{C} \to \mathcal{D}$ (which are abstract sets of objects and maps between them). (For example, one useful object we didn't explicitly state last time is the **constant functor** C_Y for some $Y \in \mathcal{D}$,

which sends any object $X \in \text{ob}(\mathcal{C})$ to Y and any $f \in \text{mor}(\mathcal{C})$ to Id_Y.) We then introduced **natural transformations** between functors $\phi : F \to G$, which associate to each $X \in \text{ob}(\mathcal{C})$ a map $\phi_X : F(X) \to G(X)$ which commutes with **any** morphism $f : X \to Y$ in $\mathcal{C}$. The point is that such a transformation is "so natural" that it doesn't depend on whether we're applying F or G. We saw an example with the determinant map $GL_n(R) \to GL_1(R)$ (where the idea was that if we have a ring morphism $R \to S$, computing the polynomial in the entires in R or in S "gives the same thing").

Example 130

Recall that we have a contravariant functor $*$ from **K-vect** to **K-vect** (where we send a vector space V to V^*), and applying $*$ twice gives us a covariant functor. From linear algebra, there is a canonical map $V \to V^{**}$ which sends any v to the evaluation map $(\lambda \mapsto \lambda(v))$, and notice that this definition didn't really depend on v in any essential way. In other words, there should be a natural transformation from the identity functor to the double dual functor.

Indeed, for any vector space $V \in$ **K-vect**, define a map $D_V : V \to V^{**}$ so that for any $v \in V$ and $\lambda \in V^*$, $D_V(v)$ is the map

$$D_V(v)(\lambda) = \lambda v.$$

We must check that the following diagram commutes for any $f : V \to W$:

$$\begin{array}{ccc} V & \xrightarrow{D_V} & V^{**} \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f^{**}} \\ W & \xrightarrow{D_W} & W^{**} \end{array}$$

To do this, we must check that starting with some $x \in V$ and going around the diagram yields the same thing in both directions, and we can do that by checking that the result is the same applied to any $\lambda \in W^*$. Indeed, we have (going around in one direction)

$$((D_W \circ f)(x))(\lambda) = D_W(f(x))(\lambda) = \lambda(f(x)) = (\lambda \circ f)(x),$$

and (in the other direction, and in the second step using the definition of the dual of a linear transformation)

$$((f^{**} \circ D_V)(x))(\lambda) = (f^{**}(D_V(x)))(\lambda) = (D_V(x) \circ f^*)(\lambda) = D_V(x)(\lambda \circ f),$$

which are indeed the same. Thus we do indeed have a natural transformation.

Example 131

The identity natural transformation from a functor F to itself sends Id_X to $\text{Id}_{F(X)}$ – we can check that this satisfies the definitions.

Example 132

If $f : Y \to Y'$ is a morphism in $\mathcal{D}$, then we can take the constant functor C_Y and get a natural transformation $C_f : C_Y \to C_{Y'}$. We must then define (for any $x \in X$) $C_{f,x} : C_Y x \to C_{Y'} x$ which basically just applies f (because $C_Y x = Y$ and $C_{Y'} x = Y'$). And this is a natural transformation because the commutative diagram is just saying that $f \circ \text{id} = \text{id} \circ f$.

Example 133
Given natural transformations $\phi : F \to G$ and $\psi : G \to H$, we get a natural transformation $(\psi \circ \phi)$ defined by $(\psi \circ \phi)_X = \psi_X \circ \phi_X$ for any X

We now may be curious about "equivalence of categories" – it turns out that requiring morphisms between them that compose to the identity is too rigid of an assumption. We'll first discuss equivalence of functors:

Definition 134
We say that two functors $F, G : \mathcal{C} \to \mathcal{D}$ are **naturally isomorphic**, denoted $F \simeq G$, if there are natural transformations $\phi : F \to G$ and $\psi : G \to F$ such that $\phi \circ \psi = \mathrm{Id}_G$ and $\psi \circ \phi = \mathrm{Id}_F$. In other words, for any $x \in \mathrm{ob}(\mathcal{C})$, we want $\phi_X : F(X) \to G(X)$ and $\psi_X : G(X) \to F(X)$ to be mutually inverse isomorphisms.

Example 135
For the category **Fin-K-vect**, the identity map and the double dual are naturally isomorphic functors (though they aren't actually equal objects, they're as good as each other for any real purpose).

Definition 136
A covariant functor $F : \mathcal{C} \to \mathcal{D}$ is an **equivalence of categories** if there is some covariant functor $G : \mathcal{D} \to \mathcal{C}$ such that $F \circ G \simeq \mathrm{Id}_{\mathcal{D}}$ and $G \circ F \simeq \mathrm{Id}_{\mathcal{C}}$. If F and G are contravariant instead and we have the same condition, we say that we have an **anti-equivalence of categories**.

Example 137
The functor $* :$ **Fin-K-vect** $\to$ **Fin-K-vect** is an anti-equivalence of categories (since we can also use $*$ to go in the opposite direction).

Definition 138
We call $\mathcal{C}$ a **small category** if $\mathrm{ob}(\mathcal{C})$ and $\mathrm{mor}(\mathcal{C})$ are both in the universe $\mathcal{U}$.

For any small category J, we can consider functors $F : J \to \mathcal{C}$, and we'll now discuss **limits** and **colimits**:

Definition 139
A **limit** (also **inverse limit**) of F, denoted $\varprojlim F$, is an element $X \in \mathrm{ob}(\mathcal{C})$ and a natural transformation $\phi : C_X \to F$ which is universal: that is, if we have any object $X' \in \mathrm{ob}(\mathcal{C})$ and natural transformation $C_{X'} \to F$, then there is a unique $\alpha : X' \to X$ such that $\psi = \phi \circ C_\alpha$.

Concretely, this means that for every $j \in \mathrm{ob}(J)$, we have a map $\phi : C_{X,j} = X \to F_j$ which commutes with all morphisms: for any $i \to j$ in J, we get the following diagram (where often we'll just collapse the two Xs into one, and the requirement is that $\phi_j = F_f \circ \phi_i$).

$$\begin{array}{ccc} X & \xrightarrow{\phi_i} & F_i \\ \downarrow{\scriptstyle \mathrm{id}} & & \downarrow{\scriptstyle F_f} \\ X & \xrightarrow{\phi_j} & F_j \end{array}$$

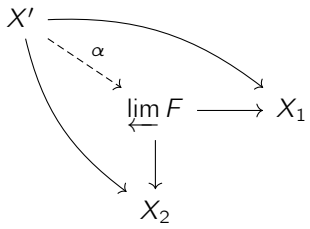
Then the universal property says that given an object X' and maps $\psi_i : X' \to F_i$ and $\psi_j : X' \to F_j$ such that $\psi_j = F_f \circ \psi_i$, we have a map $\alpha : X' \to X$ so that the following diagram commutes for all i:

$$X' \xrightarrow{\alpha} X \xrightarrow{\psi_i} F_i, \quad X' \xrightarrow{\psi_i} F_i$$

Example 140

Suppose J has two objects 1 and 2 and the two identity maps Id_1 and Id_2. Then a functor $F : J \to \mathcal{C}$ means we just need to specify that 1 goes to some $X_1 \in \mathcal{C}$ and 2 goes to some $X_2 \in \mathcal{C}$, and the identity morphisms go to the identity morphisms.

In this case, the limit $\varprojlim F$ must be some object in $\mathcal{C}$ with a map to X_1 and a map to X_2 which commute with the morphisms in J, and for any other X' with maps into X_1 and X_2 we have a unique map $X' \to \varprojlim F$:



So this actually gives us the **product** (if the limit exists – it may not exist for a general category) – more generally, for any small set I, we can create a category with objects I and only the identity morphisms. Then the limit, if it exists, is the collection of $\{X_i\}_{i \in I}$ with maps to each X_i with the product universal property that we've already seen.

Example 141

Now suppose J has two objects 1 and 2, the two identity maps Id_1 and Id_2, and two additional morphisms from 1 to 2. Then having a functor $F : J \to$ **K-vect** means that we specify two vector spaces V and W and two linear transformations $f, g : V \to W$.

The limit of this functor is then a vector space X with maps to V and W such that the result is the same whether we take f or g. In other words, it's a vector space with a map ϕ to V so that $f \circ \phi = g \circ \phi$ which also satisfies the universal property. So what we have here is the **kernel** $\ker(f - g)$ with the natural inclusion into V, and indeed for any other map $\psi : X' \to V$ with that same property, we have that $(f - g)\psi(x) = 0$ for all $x \in X'$. But that means the image of ψ lands in $\ker(f - g)$, so any such map ψ must factor through $\ker(f - g)$.

Definition 142

A **colimit** (also **direct limit**) of a functor $F : J \to \mathcal{C}$, denoted $\varinjlim F$, is an object $X \in \mathrm{ob}(\mathcal{C})$ and a natural transformation $\phi : F \to C_X$ which is universal, meaning that for any object $X' \in \mathcal{C}$ and natural transformation $\psi : F \to C_{X'}$, there is a unique $\alpha : X \to X'$ such that $\psi = C_\alpha \circ \phi$.

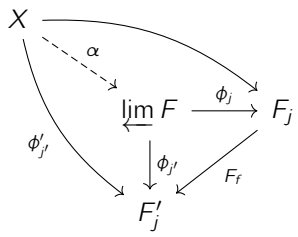
Example 143

Consider the category J of three objects $1, 2, 3$ with the identity maps and a map $1 \to 2$ and a map $1 \to 3$. Then a functor $F : J \to$ **Ring** is a specification of three rings R, S, T, along with maps $R \to S$ and $R \to T$.

Checking the definitions, a colimit $\varinjlim F$ of this functor is then the tensor product $S \otimes_R T$ (we saw the corresponding commutative diagram back when we defined the tensor product, since we need to have maps $S \to S \otimes_R T$ and $T \to S \otimes_R T$ that are compatible with the maps $R \to S$ and $R \to T$). And it turns out that limits and colimits are unique (just like we saw that the tensor product and other limit objects are unique) – we'll see that next time.

12 October 21, 2022

Last lecture, we looked at limits and colimits of functors: if $F : J \to \mathcal{C}$ is a functor, then the limit of F (which may or may not exist) are defined in the following way: we have a natural transformation $\phi : C_{\varprojlim F} \to F$ so that if $X \in \text{ob}(C)$ and we have a natural transformation $\phi' : C_X \to F$, then there is a unique morphism $\alpha : X \to \varprojlim F$ so that we have a commutative diagram $\phi' = \phi \circ C_\alpha$. Concretely, we have the following diagram for all $j \in J$, with f sending j to j':



The colimit is defined similarly. And as with many other universal properties, we have a uniqueness property:

Lemma 144

Suppose (X, ϕ) and (X', ϕ') are two limits of the functor $F : J \to \mathcal{C}$. Then there is a unique isomorphism $\alpha : X \to X'$ such that $\phi' \circ C_\alpha = \phi$ (and an analogous isomorphism $\beta : X' \to X$). Similarly, if (X, ϕ) and (X', ϕ') are two colimits of $F : J \to \mathcal{C}$, then there is a unique isomorphism $\alpha : X \to X'$ with $C_\alpha \circ \phi = \phi'$ (and an analogous isomprhism $\beta : X' \to X$).

(By the universal property, we already know that the map α uniquely exists; what we're saying is that this is an isomorphism.) Basically, we have to check that α and β are inverses in both directions, but we've done this kind of argument many times before.

Example 145

It turns out that in the category of sets, small limits and small colimits always exist – indeed, for any functor F we can define

$$\varprojlim F = \left\{ (x_j) \in \prod_{j \in \text{ob}(J)} F_j \;\middle|\; F(f)x_j = x_{j'} \ \forall f : j \to j' \right\}$$

with maps ϕ_j to F_j being projection onto the jth factor, and this is compatible with the maps $F_f : F_j \to F_{j'}$ because of the condition on the coordinates. Similarly, rings have all small limits and colimits (as we verified ourselves) – the limits are constructed similarly in the two cases, but the colimits are different (we basically have a disjoint union for sets and a tensor product for rings).

Example 146

Let J again be the category with two objects 1 and 2 and only their identity morphisms. As discussed, a functor $F : J \to \mathcal{C}$ is just a collection of two objects $(X_1, X_2) \in \text{ob}(\mathcal{C})$.

We'll see what this is more concretely for various categories.

- In **Set**, we know that the limit of F is the product set $X_1 \times X_2$, and we can check that the colimit of F is the disjoint union $X_1 \sqcup X_2$.
- In **Ring**, the limit is $X_1 \times X_2$ again, and the colimit is $X_1 \otimes_{\mathbb{Z}} X_2$ (because X_1 and X_2 map into the tensor product, and $\mathbb{Z}$ maps uniquely into any ring.)
- In **K-vect**, the limit is again the product of vector spaces, but we usually call it the direct sum $X_1 \oplus X_2$. And the colimit is again $X_1 \oplus X_2$ (because we have the map that sends $x \in X_1$ to $(x, 0)$, which is not something that works for rings); here we have the universal property because whenever we have maps $f_i : X_i \to X$, we get a map $f_1 + f_2 : X_1 \oplus X_2 \to X$ by sending (x_1, x_2) to $f_1(x_1) + f_2(x_2)$. So the point is that **K-vect** is behaving differently from **Set** and **Ring**.
- In **Grp**, the limit is again a product $X_1 \times X_2$, and the colimit is the **amalgam** $X_1 * X_2$ (because of the lack of commutativity).

Definition 147

Let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ be functors. We say that F is a **left adjoint** for G, and that G is a **right adjoint** for F, if for all $X \in \text{ob}(\mathcal{C})$ and $Y \in \text{ob}(\mathcal{D})$, there is an isomorphism $\phi_{X,Y} : \text{Hom}_{\mathcal{D}}(F(X), Y) \to \text{Hom}_{\mathcal{C}}(X, G(Y))$ which is natural in the following way: for any $f : X \to X'$, we have the commutative diagrams shown below.

For any map $f : X \to X'$, we have the following diagram ("precomposing by f," so the notation $- \circ F(f)$ means that we take some $h \in \text{Hom}_{\mathcal{D}}(F(X'), Y)$ and send it to $h \circ F(f)$):

$$\begin{array}{ccc} \text{Hom}_{\mathcal{D}}(F(X), Y) & \xrightarrow{\phi_{X,Y}} & \text{Hom}_{\mathcal{C}}(X, G(Y)) \\ {\scriptstyle -\circ F(f)}\uparrow & & \uparrow{\scriptstyle -\circ f} \\ \text{Hom}_{\mathcal{D}}(F(X'), Y) & \xrightarrow{\phi_{X',Y}} & \text{Hom}_{\mathcal{C}}(X', G(Y)) \end{array}$$

Similarly, for any map $g : Y \to Y'$, we can "postcompose by G":

$$\begin{array}{ccc} \text{Hom}_{\mathcal{D}}(F(X), Y) & \xrightarrow{\phi_{X,Y}} & \text{Hom}_{\mathcal{C}}(X, G(Y)) \\ \downarrow{\scriptstyle g\circ -} & & \downarrow{\scriptstyle G(g)\circ -} \\ \text{Hom}_{\mathcal{D}}(F(X), Y') & \xrightarrow{\phi_{X,Y'}} & \text{Hom}_{\mathcal{C}}(X, G(Y')) \end{array}$$

For another way of thinking about this definition, if $\mathcal{C}^{\text{op}}$ denotes the **opposite category** of $\mathcal{C}$, then we have a functor $\mathcal{C}^{\text{op}} \times \mathcal{D}$ to **Set** sending (x, y) to $\text{Hom}_{\mathcal{D}}(F\cdot, \cdot)$ and also one sending (x, y) to $\text{Hom}_{\mathcal{C}}(\cdot, G\cdot)$. Then ϕ is basically demonstrating a natural isomorphism between those functors.

In particular, if we take $Y = F(X)$ in our second diagram here, we have $\text{Hom}_{\mathcal{D}}(F(X), F(X))$ in the top left, which has a natural element (the identity map). Then we see that $\eta_X = \phi_{X,F(X)}(\text{Id}_{F(X)})$ sends $X \mapsto G(F(X))$, and similarly $\mu_y = \phi^{-1}_{G(Y),Y}(\text{Id}_{G(Y)})$ sends $F(G(Y)) \to Y$. So an adjunction gives a natural transformation $\eta : \text{Id}_{\mathcal{C}} \to G \circ F$ and a natural transformation $\mu : F \circ G \to \text{Id}_{\mathcal{D}}$.

It turns out that the maps η and μ determine the adjunction (they don't always give rise to an adjunction, but if they do then knowing η alone is enough). Indeed, if we have a map $\text{Hom}_{\mathcal{D}}(F(X), F(X)) \to \text{Hom}_{\mathcal{C}}(X, G(F(X))$ sending $\text{Id}_{F(X)}$ to η_X, then to complete the diagram, we need to know how to send $f \in \text{Hom}_{\mathcal{D}}$ to $\phi_{X,Y}(f) \in \text{Hom}_{\mathcal{C}}(X, G(Y))$. But that allows us to create the map in blue by composition:

$$\begin{array}{ccc} \mathrm{Hom}_{\mathcal{D}}(F(X), Y) & \longrightarrow & \mathrm{Hom}_{\mathcal{C}}(X, G(Y)) \\ {\scriptstyle g\circ -}\uparrow & & \uparrow{\scriptstyle G(g)\circ -} \\ \mathrm{Hom}_{\mathcal{D}}(F(X), F(X)) & \longrightarrow & \mathrm{Hom}_{\mathcal{C}}(X, G(F(X))) \end{array}$$

Under that map the identity map $\mathrm{Id}_{F(X)}$ goes to f, so we need to put in the red map. So determining η also allows us to construct the functor G. Similarly, if $f : X \to G(Y)$, then $\phi^{-1}_{X,Y}(f)$ is μ composed with $F(f)$, so we can determine F from the adjunction.

Example 148

Let G : **Ring** $\to$ **Set** be the forgetful map. Then G has a left adjoint – indeed, what we need is a functor F so that if Ω is a set, and R is a ring, then we have an isomorphism

$$\mathrm{Hom}_{\mathbf{Ring}}(F(\Omega), R) \cong \mathrm{Hom}_{\mathbf{Set}}(\Omega, R).$$

And what we should do is define $F(\Omega) = \mathbb{Z}[X_\omega]_{\omega\in\Omega}$ to be the **polynomial ring** (to get a map from $\mathbb{Z}[X_\omega] \to R$, we just need to designate where each X_ω goes). So we can check that $\Omega \to \mathbb{Z}[X_\omega]_{\omega\in\Omega}$ sends ω to X_ω, and $\mu_R : \mathbb{Z}[X_r]_{r\in R} \to R$ sends x_r to r.

Lemma 149

Let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ be functors so that F is left adjoint to G. Then G preserves small limits (so it sends limits to limits) and F preserves small colimits.

In particular, the forgetful functor is a right adjoint, so it preserves limits (and this explains why we saw that limits were the same in sets and rings, but not colimits). The converse turns out to be pretty true as well:

Theorem 150

Up to some set-theoretic considerations, the converse also holds: let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ be functors. Then if G preserves limits and some set theory details hold, then G has a left-adjoint. Similarly, if F preserves colimits and some set theory details hold, then F has a right adjoint.

13 October 24, 2022

Today's lecture will begin our discussion of **modules**, and we'll start with the basics:

Definition 151

Let R be a ring. An ***R*-module** M is an abelian group $(M, +)$ with a binary operation (action) $R \times M \to M$ (which will send (r, m) to rm) such that $1 \cdot m = m$ (identity), $(r + s) \cdot m = r \cdot m + s \cdot m$ and $r \cdot (m + n) = r \cdot m + r \cdot n$ (distributivity), and $r \cdot (s \cdot m) = (r \cdot s) \cdot m$ (associativity) for all $r, s \in R$ and $m, n \in M$.

(This also implies a few other immediate consequences – for example, $0 \cdot m = 0$ and $(-1) \cdot m = -m$ by plugging in appropriate constants in for m and n.)

Example 152

If R is a field, then R-modules are the same as R-vector spaces.

However, modules can be much more complicated than vector spaces, much like rings can be much more complicated than fields.

Example 153

An abelian group is the same as a $\mathbb{Z}$-module – indeed, any $\mathbb{Z}$-module is an abelian group, and for the other direction we know that $(1+1+\cdots+1)m$, where we add together n ones, must be m added to itself n times, and similarly for negatives. So the $\mathbb{Z}$-action is already specified.

Example 154

The ring R itself is an R-module; more generally, for any ideal $I \triangle R$, R/I is an R-module (in which $r(s+I) = rs+I$). And even more generally, if we have a ring homomorphism $\phi : R \to S$, then S is an R-module with $r \cdot s = \phi(r)s$ (the axioms hold because ϕ is a morphism and S is a ring).

Example 155

$\mathbb{Q}^2$ is a module over the polynomial ring $\mathbb{Q}[T]$, where (for example) we can have T act on $\begin{bmatrix} x \\ y \end{bmatrix}$ to produce $\begin{bmatrix} y \\ -x \end{bmatrix}$. More generally, any endomorphism of a K-vector space produces a $K[T]$-module, since we already know the action of K and we just need to specify what T does.

Definition 156

A map $\phi : M \to N$ is an ***R*-module morphism** (also an ***R*-linear map**) if it preserves the relevant structure, meaning that $\phi(r \cdot m + n) = r \cdot \phi(m) + \phi(n)$ for all $r \in R$ and $m, n \in M$.

In particular, we see that $\phi(0) = 0$ by setting $r = 1$ and $m = 0$.

Definition 157

A subset $N \subset M$ of an R-module is a **submodule** if N is nonempty and for all $m, n \in N$ and $r \in R$, $rm + n \in N$.

Example 158

The R-submodules of R are the ideals of R (since both definitions require being a subset and being closed under addition and multiplication by any element of R). And for any module M, (0) is a submodule.

Definition 159

We denote by **R-Mod** the category of small R-modules.

There are a few useful properties of this category:

- (0) is an initial object of this category, since it maps uniquely to everything else. Similarly, (0) is a final object. This is actually a pretty special property – in rings we had an initial object of $\mathbb{Z}$ but a final object of (0).
- The set of homomorphisms $\mathrm{Hom}_{\textbf{R-Mod}}(M, N)$ has an additive group structure, and in fact it is an R-module. (This is not something that we had in rings, because there the multiplicative structure wouldn't be preserved.) Indeed, M always has a unique map to the zero object, which has a unique map to N, so that is our identity element. Then if f, g are two homomorphisms from M to N and $r \in R$, then $rf + g$ will be a homomorphism defined by
$$(rf + g)(m) = rf(m) + g(m).$$
Further, notice that for any $m, n \in M$ and $r, s \in R$,
$$(rf + g)(sm + n) = rf(sm + n) + g(sm + n) = rsf(m) + rf(m) + sg(m) + g(n)$$
$$= s(rf(m) + g(m)) + rf(n) + g(n) = s(rf + g)(m) + (rf + g)(n),$$
so we do indeed have an R-linear map.
- We have a map $\mathrm{Hom}_{\textbf{R-mod}}(M, N) \times \mathrm{Hom}_{\textbf{R-mod}}(N, P) \to \mathrm{Hom}_{\textbf{R-mod}}(M, P)$ sending (f, g) to $g \circ f$, and this map is ***R*-bilinear**, meaning that if we hold f or g constant, the map is linear in the other (or in other words, $- \circ f : \mathrm{Hom}_{\textbf{R-mod}}(N, P) \to \mathrm{Hom}_{\textbf{R-mod}}(M, P)$ sending g to $g \circ f$ is a linear map, and so is $g \circ - : \mathrm{Hom}_{\textbf{R-mod}}(M, N) \to \mathrm{Hom}_{\textbf{R-mod}}(M, P)$ sending $f \to g \circ f$).
- We can define the direct sum of two R-modules $M \oplus N$ to be the set of ordered pairs (m, n) with $m \in M$ and $n \in N$ with addition and multiplication performed component-wise. This comes with a few natural maps: we have the R-linear map $M \hookrightarrow M \oplus N$ sending m to $(m, 0)$ and similarly $N \hookrightarrow M \oplus N$ sending n to $(0, n)$, and we also have projection $\pi_1 : M \oplus N \to M$ sending (m, n) to m and $\pi_2 : M \oplus N \to N$ sending (m, n) to n. Then we see directly that $\pi_1 \circ j_1 = \mathrm{id}_M, \pi_2 \circ j_2 = \mathrm{id}_N$, $\pi_2 \circ j_1 = 0$, $\pi_1 \circ j_2 = 0$, and finally $j_1 \circ \pi_1 + j_2 \circ \pi_2 = \mathrm{id}_{M \oplus N}$.
- It turns out that this direct sum R-module is both a product and a coproduct in **R-mod**. To verify that we have a coproduct, if we have maps $f : M \to P$ and $g : N \to P$, then there is a unique map $f + g : M \oplus N \to P$ such that $(f + g) \circ j_1 = f$ and $(f + g) \circ j_2 = g$ (sending (m, n) to $f(m) + g(n)$), and to verify that we have a product, then given any maps $f : P \to M$ and $g : P \to N$, we have a unique map $f \oplus g : P \to M \oplus N$ sending p to $(f(p), g(p))$ such that $\pi_1 \circ (f \oplus g) = f$ and $\pi_2 \circ (f \oplus g) = g$. And thus we get a canonical map between products and coproducts, which helps us recover more structure among the maps: if we have two maps $f, g : M \to N$, we can construct $(f \oplus g) : M \to N \oplus N$ and then map $\mathrm{Id} + \mathrm{Id} : N \oplus N \to N$, and the composite gives us the map $f + g$. (Similarly, we can map M into $M \oplus M$ and then map into N via $f + g$ in the coproduct sense.) So the additive structure on the homomorphisms is uniquely defined once we identify the product with the coproduct.

Definition 160

Let $f : M \to N$ be a morphism of R-modules. The **kernel** of f is the set $\ker f = \{m \in M : f(m) = 0\}$, the **image** of f is $\mathrm{im}\ f = \{f(m) : m \in M\}$, and the **cokernel** of f is $\mathrm{coker} f = N/\mathrm{im}\ f$.

It is easy to check that $\ker f$ is a submodule of M and that $\mathrm{im}\ f$ is a submodule of N.

Example 161

Let N be a submodule of M. Then the quotient abelian group M/N is an R-module by setting $r(m+N) = rm+N$, and the corresponding surjective map $\pi : M \to M/N$ has kernel N. The cokernel of $N \hookrightarrow M$ is then M/N.

We can redefine the image in terms of the kernel and cokernel, because we always have (thinking of coker f as a quotient of N)

$$\text{im } f = \ker(N \to \text{coker } f).$$

So if we know the kernel of our map, knowing about the image and cokernel are equivalent.

Example 162

If we have an injective map $f : N \to M$, then $N \cong \text{coker}(M \to \text{coker} f)$, and if we have a surjective map $N \to M$, then $M \cong \text{coker}(\ker f \to N)$. (In some sense these are the "isomorphism theorems" for R-modules.)

We'll now get into the concept of exact sequences:

Definition 163

Let $f : M \to N$ and $g : N \to P$ be two maps. Then $M \xrightarrow{f} N \xrightarrow{g} P$ is **exact** at N if im $f = \ker g$ (equivalently, $g \circ f = 0$ and if $g(n) = 0$, then $n \in \text{im } f$). A sequence $0 \to M \xrightarrow{f} N \xrightarrow{g} P \to 0$ is **short exact** if it is exact at M, N, and P (meaning that the map $M \to N$ is injective, and the map $N \to P$ is surjective, plus the same conditions as before).

Equivalently, we have a short exact sequence if M is isomorphic to $\ker g$ and coker f is isomorphic to P (we can check that the implications go both ways).

Fact 164

All of the properties we've described above make **R-mod** an **abelian category** (having direct sums that are both products and coproducts and so on). We'll see more about this when we discuss homological algebra – it turns out to be useful because we can have this notion of exactness.

We'll now talk about properties that don't hold so generally: let I be a small set and M_i an R-module for all $i \in I$. Then we can define the **direct product** (also just **product**)

$$\prod_{i \in I} M_i = \left\{ (m_i) \in \prod_{i \in I} m_i \right\}$$

and the **direct sum**

$$\bigoplus_{i \in I} M_i = \left\{ (m_i) \in \prod_{i \in I} M_i : m_i = 0 \text{ for all but finitely many } i \right\}.$$

In particular, these are only different when we get to infinitely many elements in the index set.

It turns out that **R-mod** always has small limits and colimits: if $F : J \to$ **R-mod** is a functor, then a limit of F is (basically a subset of the product)

$$\varinjlim F = \left\{ (m_j) \in \prod_{j \in \text{ob}(J)} F_j : F(\phi)(m_i) = m_j \text{ for all } \phi : i \to j \right\},$$

and similarly the colimit is (viewed as a quotient of the coproduct)

$$\varprojlim F = \bigoplus_{j \in \text{ob}(J)} F_j / \langle m_j e_j - F(\phi)(m_{j'}) e_{j'} \text{ for all } \phi : j \to j', m \in F(j) \rangle,$$

where $m_j e_j$ is shorthand for the element which has m_j in the jth component of the direct sum. Checking that we do satisfy the definitions of limits and colimits here is left as an exercise to us.

Example 165

For any small set Ω, we can form the **free module** with Ω as a basis, written as

$$F_R(\Omega) = \bigoplus_{\omega \in \Omega} R = \{F : \Omega \to R \text{ with } f(\omega) = 0 \text{ for all but finitely many } \omega\}.$$

In particular, for any $\omega \in \Omega$, we have the "basis element" element $e_\omega \in F_R(\Omega)$ so that $e_\omega(\omega') = 1$ for $\omega = \omega'$ and 0 otherwise, and then we can write an element of the free module as $(r_\omega) = \sum_{\omega \in \Omega} r_\omega e_\omega$ (if r_ω is zero except finitely often).

It turns out that F_R is left adjoint to the forgetful functor G : **R-mod** $\to$ Set. In other words, we have an isomorphism

$$\mathrm{Hom}_{\textbf{R-mod}}(F_R(\Omega), M) \cong \mathrm{Hom}_{\textbf{Set}}(\Omega, M)$$

which sends ϕ to $(\omega \mapsto \phi(e_\omega))$ and ψ to the map sending (r_ω) to $\sum_\omega r_\omega \psi(\omega)$. In other words, if ψ maps Ω to M, then there is a unique R-linear map $F_R(\Omega) \to M$ under which e_ω maps to $\psi(\omega)$. (So this is kind of like the property for polynomial rings.)

14 October 26, 2022

We'll be going over more basic constructions with modules today – last time, we mentioned the free R-module on Ω, which can be thought of as direct sums of R with itself indexed by Ω. We saw last time that this is left adjoint to the forgetful functor from **R-mod** to **Set**, so we have the universal property that specifying a map $F_R(\Omega) \to M$ is uniquely determined by specifying where each basis vector e_ω goes.

Definition 166

An R-module M is **free** if M is isomorphic to $F_R(\Omega)$ for some Ω. A subset $\mathcal{B} \subset M$ is a **basis** of M if the natural map $F_R(\mathcal{B}) \to M$ (sending each basis element e_b to b) is an isomorphism.

This definition is equivalent to saying that every element of M can be uniquely written as a **finite** R-linear combination of the elements of $\mathcal{B}$ (in other words as $\sum_{i=1}^N r_i b_i$ with $r_i \in R$ and $b_i \in \mathcal{B}$). We know that any vector space has a basis, so any R-module over a field is free. But there are definitely modules that aren't free – for example, $\mathbb{Q}$ is not a free $\mathbb{Z}$-module, because it cannot have one basis element and span all of $\mathbb{Q}$ but it cannot have at least two basis elements without having a linear relation between them. And more simply, something like $\mathbb{Z}/(2)$ is not a free $\mathbb{Z}$-module (since 0 can't be a basis element because $1 \cdot 0 = 0$, and 1 can't either because $2 \cdot 1 = 0$). The point is that being a free R-module is generally a very rare property.

It turns out that if R is nonzero and $F_R(\Omega) \cong F_R(\Omega')$, then we can put Ω and Ω' in bijection with each other. (The idea is to choose a maximal ideal and deduce that the free modules $F_{R/\mathfrak{m}}(\Omega)$ and $F_{R/\mathfrak{m}}(\Omega')$ are isomorphic, and then use the standard result for vector spaces.)

Definition 167

The **rank** of a free R-module M is the cardinality of the set Ω if $M \cong F_R(\Omega)$.

Definition 168

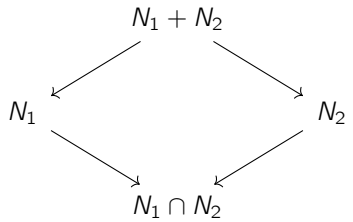
For any subset $\Omega \subset M$, the image of $F_R(\Omega)$ in M, denoted $\langle \Omega \rangle$, is the **submodule generated by** Ω.

More concretely, the submodule generated by Ω is the set of finite linear combinations $\sum_{i=1}^{n} r_i \omega_i$ with $r_i \in R$ and $\omega_i \in \Omega$. We can check that it is indeed a submodule, and whenever $N \supset \Omega$ is a submodule we also have $N \supset \langle \Omega \rangle$.

Lemma 169

For any two submodules $N_1, N_2 \subset M$, the intersection $N_1 \cap N_2$ and sum $N_1 + N_2$ are also submodules of M.

The idea is that $N_1 \cap N_2$ is the largest submodule contained in both N_1 and N_2, and $N_1 + N_2$ is the smallest submodule containing both. So we have containment much like in rings as in the diagram below:



It then turns out that we have isomorphism "along the opposite arrows:" we have

$$N_1/(N_1 \cap N_2) \cong (N_1 + N_2)/N_2,$$

defined by sending $n + (N_1 \cap N_2)$ to $n + N_2$ (as we can check).

Definition 170

Let $N \subset M$ be a submodule. Then there is a bijection between submodules of M/N and submodules of M containing N (sending a submodule P of M to P/N and sending a submodule in the quotient $\overline{P}$ to its preimage under the projection $\pi : M \to M/N$).

Definition 171

The **dual module** of M is the set of module morphisms $M^* = \mathrm{Hom}_R(M, R)$.

These dual spaces are not as well-behaved as dual spaces of vector spaces – for example, the dual module $(\mathbb{Z}/(2))^*$ is zero, because there are no nontrivial morphisms from $\mathbb{Z}/(2)$ to $\mathbb{Z}$ if twice $f(1)$ must be sent to zero. So we lose a lot of information when we pass to duals.

Definition 172

The set of endomorphisms of M is denoted $\mathrm{End}_R(M) = \mathrm{Hom}_R(M, M)$.

We saw the following as an example last time:

Lemma 173

Let M be an R-module, and let $T \in \mathrm{End}_R(M)$. Then M is also an $R[X]$ module by defining $(f_0 + f_1 x + \cdots + f_1 x^d)m = f_0 m + f_1 T m + f_2 T^2 m + \cdots + f_d T^d m$.

We'll now mention a submodule construction which does not have any nontrivial analogy for vector spaces:

Definition 174

Let $I \triangleleft R$ be an ideal. We define the multiplication $IM = \sum_{i=1}^{n} r_i m_i : r_i \in I, m_i \in M\}$.

For example, $(2)(\mathbb{Z} \oplus \mathbb{Z}) = (2) \oplus (2)$ (by checking inclusions both ways).

Definition 175

Let Ω be a subset of an R-module M. The **annihilator** of Ω in R is

$$\mathrm{Ann}_R(\Omega) = \{r \in R : r\omega = 0 \ \forall \omega \in \Omega\}.$$

We can check that this is always an ideal of R. For example, $\mathrm{Ann}_R(R/I) = I$, because anything in I multiplied by anything in R/I means we get sent to zero.

Example 176

Last time, we had an example where $\mathbb{Q}^2$ is a $\mathbb{Q}[X]$-module where X acts via the operator $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} y \\ -x \end{bmatrix}$. Then if we want the annihilator of $\mathbb{Q}^2$ in $\mathbb{Q}[X]$, then we want the set of polynomials $f \in \mathbb{Q}[X]$ such that $f(T) = 0$.

It must be an ideal of $\mathbb{Q}[X]$, so it must be principal and thus generated by some minimal polynomial (which in this case is $X^2 + 1$). Any other polynomial is then of the form $q(X)(X^2 + 1) + (aX + b)$, and this can only be in the annihilator of $aT + b = 0$, which only happens if $a, b = 0$.

Just like for rings, there are two equivalent definitions of noetherianness that are useful:

Lemma 177

Let M be an R-module. Then the following are equivalent:

1. Every submodule is finitely generated,
2. Any non-empty set of submodules contains a maximal element.

The proof here is the same as for ideals in R, and if these properties hold for an R-module M, we call M **noetherian**.

Lemma 178

We have the following properties:

1. R is noetherian as an R-module if and only if R is noetherian as a ring (because submodules of R are ideals).
2. All submodules and quotients of a noetherian module are noetherian (by property 2, since collections of submodules in each case can be thought of in terms of the submodules of the whole module).
3. As a kind of converse, if M/N and N are noetherian, then so is M.
4. If M and N are noetherian, then $M \oplus N$ is noetherian.
5. If R is noetherian and M is a finitely generated R-module, then M is noetherian.

Proof. We already gave properties of (1) and (2) above. We'll show (3) and (4) by using property 1: if P is a submodule of M, then $N \cap P$ is a submodule of N and is thus finitely generated as $\langle n_1, \cdots, n_r \rangle$ for some $n_i \in N$. But

$P/(N \cap P)$ embeds as a submodule in M/N (with the natural map which is injective) so it is also finitely generated as $\langle m_1 + N \cap P, \cdots, m_s + N \cap P \rangle$. We can then take $\langle n_1, \cdots, n_r, m_1, \cdots, m_s \rangle$ to finitely generate P, because $P - r_1 m_1 - \cdots - r_n m_n$ must be in $N \cap P$ for some $r_i \in R$ and then that is a linear combination of the n_i. And (4) follows by applying (3) to $(M \oplus N)/N \cong M$.

Finally for (5), saying that M is finitely generated says that there is some finite set Ω such that $F_R(\Omega)$ maps surjectively into M. But $F_R(\Omega)$ is the direct sum of finitely many copies of R (which is noetherian) and is thus noetherian, so M, a quotient of it, is also noetherian. □

We'll now start talking about localizations:

Definition 179

Suppose $D \subset R$ is multiplicative. We can define an equivalence relation on $M \times D$ via

$$(m, d) \sim (n, e) \text{ if } f(em - dn) = 0 \text{ for some } f \in D.$$

Then we let $D^{-1}M$ be the set of equivalence classes under this relation, and we write $\frac{m}{d} = [(m, d)]$.

We claim that $D^{-1}M$ is a $D^{-1}R$-module – indeed, we define addition and multiplication in the usual ways, with $\frac{m}{d} + \frac{n}{e} = \frac{em+dn}{ed}$, and $\frac{r}{d} \cdot \frac{m}{e} = \frac{rm}{ed}$ (and we need to check well-definedness and the axioms, but we won't do that here). Then the map $M \to D^{-1}M$ sending m to $\frac{m}{1}$ is a morphism of R-modules (note that being a $D^{-1}R$ module means we're also an R-module). We have a few other analogous definitions to the ring ones:

Definition 180

For any prime ideal $\wp \triangleleft R$, we denote $(R - \wp)^{-1}M = M_\wp$, and for any $f \in R$, we denote $\{1, f, f^2, \cdots\}^{-1}M = M_f$.

Definition 181

A submodule $N \subset M$ is **saturated** with respect to D if for any $m \in M, d \in D, dm \in N$, we must actually have $m \in N$. The ***D* saturation** of N is the submodule $\{m \in M : \exists d \in D \text{ with } dm \in N\}$.

The D saturation of N is always a submodule containing N, and it is in fact the smallest such saturated submodule. We'll do some examples next time!

15 October 28, 2022

We started discussing localization of modules last time – if $D \subset R$ is a multiplicative subset and M is an R-module, we let $D^{-1}M$ be the set of equivalence classes of $D \times M$ under the usual equivalence relation, and (as usual) we write $\frac{m}{d}$ for the equivalence class of (d, m) – we saw that this is also a $D^{-1}R$ module. We finished by defining the saturation of a submodule $N \subset M$, which is the set of $m \in M$ such that there is some d with $dm \in N$ (so "things in N divided by something in d").

Example 182

Let $R = \mathbb{Z}$ and $D = \mathbb{Z} - (3)$. Then $D^{-1}(\mathbb{Z}/(2))$ has elements of the form $\frac{a+(2)}{b}$ for some b not divisible by 3, but that element is equal to $\frac{2a+(2)}{2b} = \frac{0}{2b}$. So everything in this localization is zero and we get the trivial module.

Example 183

On the other hand, to compute $D^{-1}(\mathbb{Z}/(3))$, we know there's a $\mathbb{Z}$-linear map $\mathbb{Z}/(3) \to D^{-1}(\mathbb{Z}/3)$, and we want to see if it is injective.

We know that $a+(3)$ is sent to $\frac{a+(3)}{1}$, which cannot be zero unless $a = 0$ because that would mean $a+(3)$ is a zero divisor with some d in $\mathbb{Z} - (3)$, which is not true (since that would require 3 to divide da). So $\mathbb{Z}/(3) \to D^{-1}(\mathbb{Z}/(3))$ is injective, and it is surjective as well (because $r + (3)$ is equal to $\frac{a+(3)}{d}$ if $rd \equiv a \bmod 3$, and d is always invertible mod 3 so we can always find an r). That means $D^{-1}(\mathbb{Z}/(3)) = \mathbb{Z}/(3)$.

Example 184

The D saturation of $(3)/(6)$ as a submodule of $\mathbb{Z}/(6)$ (with the same D as before) is the set of elements $a+(6)$ such that $d(a+(6)) \in (3)/(6)$. But this is the same as saying that $3|da$, and since 3 can't divide d we must have $3|a$. So $(3)/(6)$ is already saturated.

Example 185

On the other hand, the D saturation of $(2)/(6)$ is the set of $a + (6)$ such that $2|da$ for some d not divisible by 3. But we can always just take $d = 2$, so the D saturation is all of $\mathbb{Z}/(6)$.

(To see how well this generalizes to other prime ideals, we can try this out with the example $R = \mathbb{Z}[x]$ and using the ideals (x) and (2).) We'll now list a series of localization properties (some of which we may be asked to prove). First of all, we have a universal property:

Lemma 186

For any R-module M, recall that there is a natural R-linear map $i : M \to D^{-1}M$ sending m to $\frac{m}{1}$. Then if N is a $D^{-1}R$-module and $f : M \to N$ is R-linear, then there is a unique $D^{-1}R$-linear map $\tilde{f} : D^{-1}M \to N$ such that $\tilde{f} \circ i = f$, given by

$$\tilde{f}\left(\frac{m}{d}\right) = \frac{1}{d}\tilde{f}(m).$$

(We can think of $D^{-1}M$ as the "simplest $D^{-1}R$-module" that inverts all elements of D.) If such a map $\tilde{f}$ exists, it would have to have that form above because $\tilde{f}$ needs to be $D^{-1}R$-linear, but we do have to check that the definition is well-defined. Indeed, if $\frac{m}{d} = \frac{n}{e}$ (meaning $c(em - dn) = 0$ for some $c \in D$), then we want to check if $\frac{1}{d}f(m) = \frac{1}{e}f(m)$. But because f is R-linear, we have $c(ef(m) - df(n)) = 0$, which implies that $\frac{1}{d}f(m) - \frac{1}{e}f(n) = 0$ by dividing by cde (which is in D).

Lemma 187

Let $f : M \to N$ be a morphism of R-modules. Then there is a morphism of $D^{-1}R$-modules $D^{-1}f : D^{-1}M \to D^{-1}N$ sending $\frac{m}{d}$ to $\frac{f(m)}{d}$.

Alternatively, we can construct this map with the universal property – we have maps $f : M \to N$ and the natural map $N \to D^{-1}N$ which are both R-linear, so the composition $M \to D^{-1}N$ is R-linear and thus (by the universal property) factors to a $D^{-1}R$-linear map $D^{-1}M \to D^{-1}N$. But then any $\frac{m}{d} \in D^{-1}M$ must be sent to $\frac{1}{d}\frac{f(m)}{1}$ because we can follow the images of m under both paths.

Lemma 188
We have $D^{-1}(M/N) \cong D^{-1}M/D^{-1}N$, corresponding $\frac{m+N}{d}$ with $\frac{m}{d} + D^{-1}N$.

Lemma 189
We have $D^{-1}(M \oplus N) = D^{-1}(M) \oplus D^{-1}(N)$. More generally, D^{-1} commutes with any small colimit and any finite limit (but not all limits).

For example, if we take $\prod_{i=1}^{\infty} \mathbb{Z}$ localized at the prime ideal (0), and compare that with $\prod_{i=1}^{\infty} (\mathbb{Z})_{(0)} = \prod_{i=1}^{\infty} \mathbb{Q}$. So we have a natural map from the former module to the later, sending $\frac{(m_i)}{d} \to \left(\frac{m_i}{d}\right)$, but it's not an isomorphism (not surjective) because we have bounded denominators on the left-hand side but not on the right-hand side.

Lemma 190
The localization of a free module is given by $D^{-1}F_R(\Omega) = F_{D^{-1}R}(\Omega)$.

Lemma 191
For any ideal I of R, we have $D^{-1}(IM) \cong (D^{-1}I)(D^{-1}M)$.

Lemma 192
If we have a sequence $M \xrightarrow{f} N \xrightarrow{g} P$ which is exact at N, then $D^{-1}M \xrightarrow{D^{-1}f} D^{-1}N \xrightarrow{D^{-1}g} D^{-1}P$ is exact at $D^{-1}N$. Furthermore, if $0 \to M \to N \to P \to 0$ is short exact, then so is $0 \to D^{-1}M \to D^{-1}N \to D^{-1}P \to 0$.

Lemma 193
Localization preserves kernels and cokernels: if $f : M \to N$ is a module morphism, then $\ker(D^{-1}f) = D^{-1}\ker(f)$ and $\operatorname{coker}(D^{-1}f) = D^{-1}(\operatorname{coker} f)$.

Lemma 194
Let $i : M \to D^{-1}M$ be the natural map. Then we can map between (1) submodules of $D^{-1}M$, (2) the D-saturated submodules of M, and (3) all submodules of M in the following ways:

- (1) $\to$ (2): we can pull back a submodule of $D^{-1}M$ via i^{-1} and always get a D-saturated submodule of M.
- (3) $\to$ (1): since localization preserves injections, we can send a submodule N to $D^{-1}N$
- We can map (2) $\to$ (3) via inclusion and (3) $\to$ (2) via D saturation.

Then the diagram commutes (that is, applying (3) $\to$ (1) $\to$ (2) is the same as saturation, the maps (2) $\to$ (3) $\to$ (2) give us the identity map, and this shows (1) $\to$ (2) is a bijection.

All of the properties so far should be fairly routine to check, but this next one is a bit more difficult:

Lemma 195
If R is noetherian and M is finitely generated over R, then

$$D^{-1}\mathrm{Hom}_{R\text{-}\mathbf{mod}}(M, N) \cong \mathrm{Hom}_{D^{-1}R\text{-}\mathbf{mod}}(D^{-1}M, D^{-1}N).$$

Example 196

Let $R = \mathbb{Z}$ and $D = \mathbb{Z} - (3)$. Then $D^{-1}(\mathbb{Z}/(6)) = D^{-1}(\mathbb{Z}/(2)) \oplus D^{-1}(\mathbb{Z}/(3))$ (as rings we'd say $\times$ but here we'll use direct sum), and by functoriality this is $D^{-1}(\mathbb{Z}/(2)) \oplus D^{-1}(\mathbb{Z}/(3)) \cong \mathbb{Z}/(3)$ from our earlier calculations. We can check also that $D^{-1}(\mathbb{Z}/(9)) = \mathbb{Z}/(9)$.

We've mentioned that $D^{-1}M$ is the "simplest $D^{-1}R$-module," and motivated by that, let $\phi : R \to S$ be any ring morphism. We want a way to get an S-module from an R-module, and for that we'll make the following definition:

Definition 197

Let $\phi : R \to S$ be a ring morphism, and let M be an R-module. Then define the tensor product $S \otimes_R M = S \otimes_{\phi,R} M$ to be

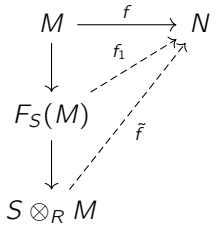
$$S \otimes_R M = F_S(M)/\langle e_n + e_m - e_{n+m}, e_{rn} - \phi(r)e_n : m, n \in M, r \in R\rangle.$$

By definition this is an S-module (since we have a quotient of a free S-module) – any element of $S \otimes_R M$ is a finite sum of the pure tensors $s \otimes m$, and we have the properties $t(s \otimes m) = (ts) \otimes m$, $s \otimes (m + n) = s \otimes m + s \otimes n$, $(s + t) \otimes m = s \otimes m + t \otimes m$, and $s \otimes (rm) = s\phi(r) \otimes m$. Furthermore, we have a natural map $M \to S \otimes_R M$ sending $m \mapsto 1 \otimes m$. But what's difficult (just like with tensor products of rings) is that we don't know what all the relations are because we can have complicated linear combinations. Instead, we work with the universal property:

Lemma 198

Let N be any S-module. Then if $f : M \to N$ is R-linear (meaning that $f(rm_1 + m_2) = \phi(r)f(m_1) + f(m_2)$ for all $m_1, m_2 \in M$ and $r \in R$), then there is a unique S-linear map $\tilde{f} : S \otimes_R M \to N$ such that f is $\tilde{f}$ composed with the natural map $M \to S \otimes_R M$ – in particular, we must have $\tilde{f}(s \otimes m) = \tilde{f}(s(1 \otimes m)) = sf(m)$ on the pure tensors.

The only question we must ask is existence, and we do that by constructing the following diagram:



As sets, we know that M embeds into $F_S(M)$, so by the universal property of the free module we can construct the map f_1 given the map f. Then we get $\tilde{f}$ from f_1 by the universal property of the quotient, because everything in the kernel $F_S(M) \to S \otimes_R M$ does get sent to zero (indeed, $f_1(e_n + e_m - e_{n+m}) = f(n) + f(m) - f(n + m) = 0$ and similar for the other generators). Next time, we'll see that if M is actually a ring, then this coincides with the definition before of tensor products of rings!

16 October 31, 2022

Last lecture, we introduced tensor products of an R-module with a ring: specifically, if $\phi : R \to S$ is a ring morphism and M is an R-module, we defined the S-module $S \otimes_R M$, consisting of finite sums of pure tensors $s \otimes m$. In this tensor product, we turn out to have $t(s \otimes m) = ts \otimes m$, linearity in each argument, and $s \otimes rm = \phi(r)(s \otimes m)(\phi(r)s) \otimes m$; we

also have a natural R-linear map $M \to S \otimes_R M$ sending $m \to 1 \otimes m$. And the key universal property here is that for any S-module N and any R-linear map $f : M \to N$ (with respect to ϕ, meaning that $f(rm_1 + m_2) = \phi(r)f(m_1) + f(m_2)$), we get a unique map $\tilde{f} : S \otimes_R M \to N$ (sending $s \otimes m$ to $f(s) \otimes m$) such that f factors through $S \otimes_R M$.

We have that $S \otimes_R -$ is left adjoint to the **forgetful functor** $\phi^* :$ **S-mod** $\to$ **R-mod**, where for any S-module N this forgetful functor preserves the abelian group structure and $r \cdot n = \phi(r)n$. Then we have

$$\mathrm{Hom}_S(S \otimes_R M, N) \cong \mathrm{Hom}_R(M, N).$$

We'll state some basic properties of this tensor product mostly without proof:

Lemma 199

The following tensor product constructions are all equivalent for any rings R, S, T and M an R-module:

1. $S \otimes_R R \cong S$,
2. $R/I \otimes_R M \cong M/IM$,
3. $(D^{-1}R) \otimes_R M \cong D^{-1}M$,
4. $T \otimes_S (S \otimes_R M) \cong T \otimes_R M$,
5. $S \otimes_R (M \oplus N) = (S \otimes_R M) \oplus (S \otimes_R N)$ (because left adjoints always preserve coproducts and the direct sum is both a product and a coproduct – this is the sort of thing that adjoints are good for),
6. For any set Ω, $S \otimes_R F_R(\Omega) = F_S(\Omega)$ (again this is because we really have a coproduct).

The first four of these basically follow by checking that we have the same universal property for the left and right sides.

Lemma 200

If $f : M \to N$ is a morphism of R-modules, then there is a unique morphism $1 \otimes f : S \otimes_R M \to S \otimes_R N$ sending $s \otimes m \to s \otimes f(m)$.

To show existence, we want a map from $S \otimes_R M \to S \otimes_R N$, which we construct by first constructing a map $M \to S \otimes_R N$ sending m to $1 \otimes f(m)$. (So we're plugging in $S \otimes_R N$ into the universal property as N.) This is manifestly well-defined, and it's linear because

$$1 \otimes f(rm_1 + m_2) = 1 \otimes (rf(m_1) + f(m_2)) = \phi(r)(1 \otimes f(m_1)) + (1 \otimes f(m_2))$$

by linearity of f and the properties of pure tensors. Thus we must also get a unique map $S \otimes_R M \to S \otimes_R N$.

Lemma 201

Let $\phi : R \to S$ and $\psi : R \to T$ be ring morphisms. Then the **ring** tensor product $S \otimes_R^{\mathrm{ring}} T$ is isomorphic (as an S-module) to the **module** tensor product $S \otimes_R^{\mathrm{mod}} T$ that we've defined, sending $s \otimes t$ to $s \otimes t$.

(It's slightly remarkable that our constructions are the same – in one case, we looked at the free module over elements of T and modded out by relations, and in the other we looked at the polynomial ring over elements of both S and T and modded out by an ideal.)

Proof. To construct a map $S \otimes_R^{\text{mod}} T \to S \otimes_R^{\text{ring}} T$, consider the map $T \to S \otimes_R^{\text{ring}} T$ sending t to $1 \otimes t$. This map is R-linear because $rt_1 + t_2$ is sent to $1 \otimes (t_1 + rt_2) = (1 \otimes t_1) + r(1 \otimes t_2)$, so it extends to an S-linear map $S \otimes_R^{\text{mod}} T \to S \otimes_R^{\text{ring}} T$ sending $s \otimes t$ to $s(1 \otimes t) = s \otimes t$, as desired.

Checking injectivity and surjectivity is generally very hard, so we usually want to construct an inverse map. But for the other direction the trouble is that we don't even know that $S \otimes_R^{\text{mod}} T$ is a ring – it's an S-module, but we haven't defined any multiplication on it yet. So we should establish a ring structure first. Given any $x \in S \otimes_R^{\text{mod}} T$, we need to define a "multiplication-by-x" map $m(x) : S \otimes_R^{\text{mod}} T \to S \otimes_R^{\text{mod}} T$ (so $m(x) \in \text{End}_S(S \otimes_R^{\text{mod}} T)$), and now this is a little easier to deal with because we know what linear maps look like even if we don't strictly have "multiplication."

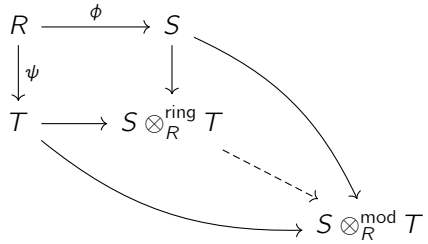
But we also know that we want $m(s \otimes t)$ to be sent to the map $(u \otimes v \mapsto su \otimes tv)$ (that's what multiplication by $s \otimes t$ does), and now we're in the realm where we can use the universal property: we have a map $T \to \text{End}_R(T)$ (sending t to "multiply by t"), which then maps into $\text{End}_S(S \otimes_R^{\text{mod}} T)$ (extending by S-linearity, sending f to $1 \otimes f$). The composite of these maps $m_1 : T \to \text{End}_S(S \otimes_R^{\text{mod}} T)$ sends t to the map $(u \otimes v \mapsto u \otimes tv)$, and this is R-linear, because $u \otimes (rt_1 + t_2)v = r(u \otimes t_1 v) + (u \otimes t_2 v)$ as usual by the basic properties of tensors. Thus by the universal property of tensor products (of modules), we get a map $S \otimes_R T \to \text{End}_S(S \otimes_R^{\text{mod}} (T)$, which we call m, which sends $s \otimes t$ to the map $m(s \otimes t)$ such that our composite map m_1 factors through it, meaning

$$m(s \otimes t)(u \otimes v) = s(m_1 t)(u \otimes v) = s(u \otimes tv) = su \otimes tv.$$

So we do have a unique S-linear map m, and now we can define a product $(S \otimes_R^{\text{mod}} T) \times (S \otimes_R^{\text{mod}} T) \to S \otimes_R^{\text{mod}} T$ sending $(x, y) \to m(x)y$ – we claim this is the multiplication structure that we want to make $S \otimes_R^{\text{mod}} T$ into a ring.

- To check distributivity, we must check that $m(x)(y + z) = m(x)y + m(x)z$, but this is true by linearity of $m(x)$. Similarly, we must check that $m(x + y)(z) = (m(x) + m(y))z = m(x)z + m(y)z$, which is true by linearity of m itself.
- To check commutativity, we will make use of distributivity – any x is a finite sum of pure tensors, so we can just check that $m(x)y = m(y)x$ when x and y are pure tensors. But we already have a formula $m(s \otimes t)(u \otimes v) = su \otimes tv$, and we have commutativity in S and T so this is the same as $m(u \otimes v)(s \otimes t)$.
- Similarly, we can reduce associativity to pure tensors (where the result is clear).
- The identity multiplication element is $1 \otimes 1$, and we can check that it is actually an identity by reducing to pure tensors.

So we've constructed a natural multiplication map on $S \otimes_R^{\text{mod}} T$, and to actually get our map from the ring tensor product to the module tensor product, we need ring morphisms from S and T into $S \otimes_R^{\text{mod}} T$. We map $T \to S \otimes_R^{\text{mod}} T$ by sending 1 to $1 \otimes t$ – this is a morphism of rings because $t_1 + t_2$ is indeed sent to $1 \otimes (t_1 + t_2)$ and $t_1 t_2$ is sent to $1 \otimes (t_1 t_2) = (1 \cdot 1) \otimes (t_1 \cdot t_2) = (1 \otimes t_1)(1 \otimes t_2)$ (by our formula for multiplication on $S \otimes_R^{\text{mod}} T$ for pure tensors). Similarly we map $S \to S \otimes_R^{\text{mod}} T$ by sending s onto $s \otimes 1$. So now we have the familiar tensor product of rings diagram:



The outer diagram commutes because going around the top sends r to $\phi(r)$ to $\phi(r) \otimes 1$, and going around the bottom sends r to $\psi(r)$ to $1 \otimes \psi(r)$. But the point is that linearity gives $1 \otimes \psi(r) = 1 \otimes \psi(r) \cdot 1 = \phi(r) \otimes 1$ in the

module tensor product. Thus we get a unique dashed map which must send $s \otimes t$ to $(s \otimes 1)(1 \otimes t) = s \otimes t$. These maps are clearly mutual isomorphisms, as desired. □

The point is that we actually can do more complicated arguments with universal properties, and not all proofs with them will be short!

Next time, we'll look at a way to construct the tensor products of two general R-modules, and this will require us to think about **multilinear algebra**:

Definition 202

Let $M_1, \cdots, M_n, N$ be R-modules. A map $\phi : M_1 \times \cdots \times M_n \to N$ (where we do not think of the left-hand side as a direct sum of R-modules, just as a tuple) is **multilinear** if it is linear if we fix the arguments for all but one M_i. In other words, ϕ is multilinear if $\phi(m_1, \cdots, m_{i-1}, rm_i + m_i', m_{i+1}, \cdots, m_n) = r\phi(m_1, \cdots, m_{i-1}, m_i, m_{i+1}, \cdots, m_n) + \phi(m_1, \cdots, m_{i-1}, m_i', m_{i+1}, \cdots, m_n)$ for all $r \in R$, all $1 \leq i \leq n$, and all $m_i, m_i' \in M_i$.

For example, if $R = \mathbb{R}$ and $M = \mathbb{R}^3$, then the map $M \times M \to \mathbb{R}$ sending (x, y) to the dot product $x \cdot y$ is bilinear (multilinear for two variables), and the map $M \times M \to M$ sending (x, y) to the cross product $x \times y$ is also bilinear. But we'll get into it more next time.

17 November 2, 2022

We defined **multilinear maps** last time – given R-modules $M_1, \cdots, M_a, P$, a multilinear map is a map from the set-theoretic product $\psi : M_1 \times \cdots \times M_a \to P$ which is linear in each variable when we fix all other variables. The usual dot (inner) product and cross (vector) product from vector calculus are multilinear maps, and we'll start today by mentioning some other important properties:

Definition 203

A multilinear map ψ is **symmetric** if $M_1 = \cdots = M_a = M$ (all modules are the same) and for any permutation $\sigma \in S_a$, we have $\psi(m_1, \cdots, m_a) = \psi(m_{\sigma(1)}, \cdots, m_{\sigma(a)})$.

In other words, it doesn't matter what order we input our arguments in – the dot product is symmetric but not the cross product. The cross product instead falls into another category of linear maps:

Definition 204

A multilinear map ψ is **alternating** if $M_1 = \cdots = M_a = M$ and whenever $m_i = m_j$ for some $i \neq j$, $\psi(m_1, \cdots, m_a) = 0$.

We can restate this definition to look more similar to the symmetric case:

Lemma 205

If ψ is alternating and $\sigma \in S_a$ is any permutation, then $\psi(m_{\sigma(1)}, \cdots, m_{\sigma(a)}) = \operatorname{sgn}(\sigma)\psi(m_1, \cdots, m_a)$ (this is sometimes called being **antisymmetric**). The converse also holds if $2 \in R^\times$.

Proof. Since the permutation group is generated by transpositions, it's sufficient to check the case where σ is a transposition. To simplify the notation, we'll just consider the transposition (12), but the argument is the same in all

cases. By multilinearity and the definition of being alternating,

$$0 = \psi(m_1 + m_2, m_1 + m_2, m_3, \cdots, m_a) = \psi(m_1, m_1, m_3, \cdots, m_a) + \psi(m_1, m_2, m_3, \cdots, m_a) + \psi(m_2, m_1, m_3, \cdots, m_a) + \psi(m_2, m_2, m_3, \cdots, m_a).$$

But the first and last terms on the right-hand side are zero by the alternating property again, so $\psi(m_1, m_2, m_3, \cdots, m_a) = -\psi(m_2, m_1, m_3, \cdots, m_a)$, as desired (because any transposition has sign -1). And for the converse, if $\psi(m_1, m_1, m_3, \cdots, m_a) = 0$, then by the antisymmetric property we have $\psi(m_1, m_1, m_3, \cdots, m_a) = -\psi(m_1, m_1, m_3, \cdots, m_a)$, so combining terms and dividing by 2 proves the result. □

We will denote the set of bilinear maps (multilinear maps with two inputs) $M_1 \times M_2 \to P$ by $\text{Bil}_R(M_1 \times M_2, P)$, and we could do something similar for the set of multilinear maps. This is not only a set but also naturally an R-module, since $(r\psi + \phi)$ is a bilinear map if ψ and ϕ are by setting

$$(r\psi + \phi)(m_1, m_2) = r\psi(m_1, m_2) + \phi(m_1, m_2).$$

Lemma 206

There is a natural isomorphism $\text{Bil}_R(M_1 \times M_2, P) \cong \text{Hom}_R(M_1, \text{Hom}_R(M_2, P))$ sending ψ to $(m_1 \to (m_2 \to \psi(m_1, m_2)))$ in the forward direction and sending any f to $((m_1, m_2) \mapsto f(m_1)(m_2))$.

We can check that both of these are R-linear maps, they do indeed end up being module morphisms / bilinear maps in the corresponding directions, and they are inverses of each other. It turns out all of this helps us define the tensor product of two R-modules:

Proposition 207

There is a universal multilinear map from $M_1 \times \cdots \times M_a$ to an R-module that we will denote $M_1 \otimes \cdots \otimes M_a$ (called the **tensor product** of $M_1, \cdots, M_a$), sending $(m_1, \cdots, m_a) \to m_1 \otimes \cdots \otimes m_a$, so that if $\psi : M_1 \times \cdots \times M_a \to P$ is any multilinear map, then there is a unique R-linear (not multilinear) map $\tilde{\psi} : M_1 \otimes \cdots \otimes M_a \to P$ such that ψ factors through $M_1 \otimes \cdots \otimes M_a$ via $\tilde{\psi}$.

Proof. We use the usual construction: consider the free R-module $F_R(M_1 \times \cdots \times M_a)$ with one generator for each element of the set-theoretic product. Then any function $M_1 \times \cdot \times M_a \to P$ factors through $F_R(M_1 \times \cdots \times M_a)$, but we need some additional constraints if we want this to be true for only multilinear maps. Thus, we must mod out by some relations, so we will define

$$F_R(M_1 \times \cdots \times M_a) / \left\langle e_{(m_1, \cdots, m_i + rm_i', \cdots, m_a)} - e_{(m_1, \cdots, m_i, \cdots, m_a)} - re_{(m_1, \cdots, m_i', \cdots, m_a)} \ \forall i, m_i' \in M_i, m_j \in M_j, r \in R \right\rangle$$

$F_R(M_1 \times \cdots \times M_a)$ surjects onto $M_1 \otimes \cdots \otimes M_a$, and for any multilinear map $M_1 \times \cdots \times M_a \to P$ everything in the kernel that we modded out by above gets sent to zero under ψ. Thus we do indeed get a unique map (by universal property of the quotient) $M_1 \otimes \cdots \otimes M_a \to P$. □

With this construction, we can define the **pure tensors**

$$m_1 \otimes \cdots \otimes m_a = [e_{(m_1, \cdots, m_a)}],$$

which span $M_1 \otimes \cdots \otimes M_a$, and we see that we must actually have $\tilde{\psi}$ defined as

$$\tilde{\psi}(m_1 \otimes \cdots \otimes m_a) = \psi(m_1, \cdots, m_a).$$

Also, we see immediately from the relations that we have

$$m_1 \otimes \cdots \otimes (m_i + rm_i') \otimes \cdots \otimes m_a = m_1 \otimes \cdots \otimes m_i \otimes \cdots \otimes m_a + rm_1 \otimes \cdots \otimes m_i' \otimes \cdots \otimes m_a.$$

But as usual, we don't want to work with the actual construction of the tensor product – it's possible to map into the tensor product, but it's hard to define maps out of the tensor product from the generators because there are lots of relations that are hard to check.

Lemma 208

There is a universal symmetric multilinear map $M \times \cdots \times M \to S^a(M)$ (this latter R-module is called the **symmetric power**) such that any symmetric multilinear map $\psi : M \times \cdots \times M \to P$ factors through a map $\tilde{\psi} : S^a(M) \to P$.

Basically the same proof works – we look at the free module $F_R(M \times \cdots \times M)$ and mod out by the relations above, but also mod out by $e_{(m_1,\cdots,m_a)} - e_{(m_{\sigma(1)},\cdots,m_{\sigma(a)})}$. And we'll often just write the elements of $S^a(M)$ as $m_1 \otimes \cdots \otimes m_a$ like the tensor product.

Lemma 209

There is a universal alternating multilinear map $M \times \cdots \times M \to \Lambda^a(M)$ such that any alternating multilinear map $\psi : M \times \cdots \times M \to P$ factors through a map $\tilde{\psi} : \Lambda^a(M) \to P$.

Again we just write down similar relations as before, and the notation we use is that $(m_1, \cdots, m_a)$ is sent to $m_1 \wedge \cdots \wedge m_a$ in $\Lambda^a(M)$.

Lemma 210

Suppose that $f_i : M_i \to N_i$ is R-linear. Then there is a unique R-linear map $f_1 \otimes \cdots \otimes f_a : M_1 \otimes \cdots \otimes M_a \to N_1 \otimes \cdots \otimes N_a$ such that $m_1 \otimes \cdots \otimes m_a$ is mapped to $f_1(m_1) \otimes \cdots \otimes f_a(m_a)$.

Proof. Such a map must be unique if it exists (since we define it on all pure tensors and can extend by linearity). To show existence, we first define a multilinear map $\psi : M_1 \times \cdots \times M_a \to N_1 \otimes \cdots \otimes N_a$ sending $(m_1, \cdots, m_a)$ to $f_1(m_1) \otimes \cdots \otimes f_a(m_a)$. This is indeed R-multilinear because

$$\begin{aligned}
\psi(m_1, \cdots, m_i + rm_i', \cdots, m_a) &= f_1(m_1) \otimes \cdots \otimes f_i(m_i + rm_i') \otimes \cdots \otimes f_a(m_a) \\
&= f_1(m_1) \otimes \cdots \otimes (f_i(m_i) + rf_i(m_i')) \otimes \cdots \otimes f_a(m_a) \\
&= f_1(m_1) \otimes \cdots \otimes f_i(m_i) \otimes \cdots \otimes f_a(m_a) + rf_1(m_1) \otimes \cdots \otimes f_i(m_i') \otimes \cdots \otimes f_a(m_a)
\end{aligned}$$

by R-linearity of f_i. Thus by the universal property of the tensor product we do indeed get $f_1 \otimes \cdots \otimes f_a : M_1 \otimes \cdots \otimes M_a \to N_1 \otimes \cdots \otimes N_a$, doing the right thing to pure tensors. □

Similar results hold for the subclasses of multilinear maps as well:

Lemma 211

If $f : M \to N$ is R-linear, then there is a unique R-linear map $S^a(f) : S^a(M) \to S^a(N)$ sending $m_1 \otimes \cdots \otimes m_a$ to $f(m_1) \otimes \cdots \otimes f(m_a)$. Similarly, there is a unique R-linear map $\Lambda^a(f) : \Lambda^a(M) \to \Lambda^a(N)$ sending $m_1 \wedge \cdots m_a$ to $f(m_1) \wedge \cdots \wedge f(m_a)$.

We prove these facts similarly by constructing a map starting with $M \times \cdots \times M$ and use the fact that we map into $S^a(N)$ or $\Lambda^a(N)$ to verify the additional requirement for the universal properties.

Remark 212. *Sometimes tensor products behave in a funny way and we should be careful: for example, take the ideal (X, Y) of $\mathbb{C}[X, Y]$, which is* **torsion-free** *(meaning that no non-zero divisor of the ring kills any element of the ideal), but $(X, Y) \otimes_R (X, Y)$ turns out to have torsion. So it takes some time to get the right intuition for all of this.*

Lemma 213

Let $\sigma \in S_a$. Then there is an R-linear isomorphism $\sigma^* : M_1 \otimes \cdots \otimes M_a \to M_{\sigma(1)} \otimes \cdots \otimes M_{\sigma(a)}$ sending $m_1 \otimes \cdots \otimes m_a$ to $m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(a)}$.

Proof. As usual, first write down a multilinear map $M_1 \times \cdots \times M_a \to M_{\sigma(1)} \otimes \cdots \otimes M_{\sigma(a)}$ sending $(m_1, \cdots, m_a) \mapsto m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(1)}$; this is multilinear so it factors through the tensor product. To show this is an isomorphism, we can also check that $(\sigma\tau)^* = \tau^*\sigma^*$ by checking that the maps agree on pure tensors (we should be careful about the reversing of order and check it carefully), so $\mathrm{id} = \mathrm{id}^* = (\sigma \circ \sigma^{-1})^* = (\sigma^{-1})^* \circ \sigma^*$ (and the same in the other direction) so we do have an inverse map. □

Lemma 214

For any $a > b$, we have

$$(M_1 \otimes \cdots \otimes M_b) \otimes (M_{b+1} \otimes \cdots \otimes M_a) \cong M_1 \otimes \cdots \otimes M_a.$$

Proof to be continued. The forward direction is a bit tricky, because we need to construct a bilinear map and the universal property doesn't tell us anything about how to construct such maps. So we start with the reverse direction: we can construct a map $M_1 \times \cdots \times M_a \to (M_1 \otimes \cdots \otimes M_b) \otimes (M_{b+1} \otimes \cdots \otimes M_a)$ sending $(m_1, \cdots, m_a)$ to $(m_1 \otimes \cdots \otimes m_b) \otimes (m_{b+1} \otimes \cdots \otimes m_a)$. This is indeed a multilinear map, which is true by the multilinearity of tensor products twice (on the inner one, then the outer one), so we get the desired map in the reverse direction. We'll do the other direction next time. □

18 November 4, 2022

We discussed properties of tensor products of modules last time – we'll go through material a bit faster than usual today because we're a bit behind. We started showing last time that the tensor product $(M_1 \otimes \cdots \otimes M_a)$ is isomorphic to the tensor product of $(M_1 \otimes \cdots \otimes M_b)$ and $(M_{b+1} \otimes \cdots \otimes M_a)$ (sending pure tensors in the way that is visually clear – $m_1 \otimes \cdots \otimes m_a$ is sent to $(m_1 \otimes \cdots \otimes m_b) \otimes (m_{b+1} \otimes \cdots \otimes m_a$ and vice versa). Starting from $(M_1 \otimes \cdots \otimes M_a)$ and constructing a map to the other space is easy, because we can map from $M_1 \times \cdots \times M_a$ into $(M_1 \otimes \cdots \otimes M_b)$ and $(M_{b+1} \otimes \cdots \otimes M_a)$ and show that it is R-linear. But the proof in the other direction is slightly more complicated:

Proof. To construct a map $(M_1 \otimes \cdots \otimes M_b) \times (M_{b+1} \otimes \cdots \otimes M_a)$ to $M_1 \otimes \cdots \otimes M_a$, we need that map to be **bilinear**. Recall that we have the isomorphism $\mathrm{Bil}_R(M_1 \times M_2, P) \cong \mathrm{Hom}_R(M_1, \mathrm{Hom}_R(M_2, P))$, so we want to construct a map $\mathrm{Hom}(M_1 \otimes \cdots \otimes M_b, \mathrm{Hom}(M_{b+1} \otimes \cdots \otimes M_a, M_1 \otimes \cdots \otimes M_a))$. Indeed, the construction should be

$$m_1 \otimes \cdots \otimes m_b \mapsto (m_{b+1} \otimes \cdots \otimes m_a \mapsto m_1 \otimes \cdots \otimes m_a).$$

We must check that for any fixed $m_1, \cdots, m_b$, we do indeed have such a morphism of R-modules. But we know that the map from $M_1 \times \cdots \times M_b$ (the product, not tensor product) to $\text{Hom}(M_{b+1} \otimes \cdots \otimes M_a, M_1 \otimes \cdots \otimes M_a)$ is an R-multilinear map, so it factors through the tensor product and thus we do have a linear map $f_{(m_1,\cdots,m_b)}$ for each $m_1, \cdots, m_b$ which is well-defined. (In particular, this tells us that even if we have a weird combination of pure tensors which is actually zero, it will get sent to zero.) So now the map $f : M_1 \times \cdots \times M_b$ (again product, not tensor product) to $\text{Hom}(M_{b+1} \otimes \cdots \otimes M_a, M_1 \otimes \cdots \otimes M_a)$ sending $(m_1, \cdots, m_b) \to f_{(m_1,\cdots,m_b)}$ is multilinear (because $f_{(rm_1+m_1',m_2,\cdots,m_b)}$ and $rf_{(m_1,\cdots,m_b)} + f_{(m_1',m_2,\cdots,m_b)}$ evaluate to the same thing on all pure tensors by multilinearity of the tensor product in each entry). Thus f also factors through the tensor product to a map $\tilde{f}$ such that $\tilde{f}(m_1 \otimes \cdots \otimes m_b) = (m_{b+1} \otimes \cdots \otimes m_a)$.

To make that into a bilinear form, we define $\phi(x, y) = \tilde{f}(x)(y)$. Such a bilinear form is a bilinear map $(M_1 \otimes \cdots \otimes M_b) \times (M_{b+1} \otimes \cdots \otimes M_a)$ to $M_1 \otimes \cdots \otimes M_a$, meaning it factors to the tensor product. So we have the map in the other direction as what we constructed last time, and it does so in a way that makes the two maps inverses (because it is on all pure tensors). Thus we have the desired isomorphism. □

We'll now go through some other properties:

Lemma 215

Let $\psi : R \to S$ be a ring morphism and M be an R-module. Then the tensor product $S \otimes_R^{\text{mod}} M$ can be thought of as a getting an S-module from an R-module under the action of S (the ring-module tensor product we first defined, giving us an S-module and thus an R-module via ψ), or as an R-module $S \otimes_R^{\text{bil}} M$ (where we think of both S and M as R-modules to start with and then do our bilinear map construction). But these constructions are isomorphic by sending $s \otimes m$ to $s \otimes m$ in both directions.

Much like in the previous argument, going backwards is easy, but going forward is tricky because we first have to show that the R-module is actually an S-module. So we have to explain how multiplication by an element of S actually looks before we can invoke the universal property.

Lemma 216

If M, N, P are all R-modules, then $M \otimes (N \oplus P) \cong (M \otimes N) \oplus (M \otimes P)$, sending $m \otimes (n, p)$ to $(m \otimes n, m \otimes p)$ and vice versa.

(This turns out to be true for arbitrary direct sums, not just finite ones.)

Lemma 217

As a special case, we have $F_R(X) \otimes F_R(Y) \cong F_R(X \times Y)$, sending $e_x \otimes e_y$ to $e_{(x,y)}$ and vice versa.

Lemma 218

The ath symmetric power $S^a(F_R(X))$ is isomorphic to the free module $F_R(S^a(X))$, where $S^a(X)$ is the a-fold product $X \times \cdots \times X$ quotiented out by the action of the symmetric group S_a (in other words, multisets of a elements), where we get an isomorphism via $e_{x_1} \otimes \cdots \otimes e_{x_a}$ to $e_{[(x_1,\cdots,x_a)]}$.

So on free modules, tensor products behave quite well – it's only when we get to arbitrary tensor products that things get messier.

The corresponding ath exterior power $\Lambda^a(F_R(X))$ can be described in a more complicated way: let $<$ be a total order on X (which always exists but is not unique). Then we define

$$\Lambda^a_<(X) = \{(x_1, \cdots, x_a) : x_i \in X, x_1 < \cdots < x_a\},$$

Lemma 219

We have an isomorphism $\Lambda^a(F_R(X)) \cong F_R(\Lambda^a_<(X))$, sending $e_{x_1} \wedge \cdots \wedge e_{x_a}$ to 0 if $x_i = x_j$ for some $i \neq j$ and $\operatorname{sgn}(\sigma) e_{(x_{\sigma(1)}, \cdots, x_{\sigma(a)})}$ if σ is the permutation that puts the x_is in increasing order.

For example, if $R^{\oplus a}$ is the direct sum of a copies of R, which is a free module on $\{1, \cdots, a\}$, then there is only one element of $\Lambda^a_<(X)$, so

$$\Lambda^a(R^{\oplus a}) \cong R,$$

though this isomorphism is not canonical. Similarly, we have $\Lambda^{a-1}(R^{\oplus a}) \cong R^{\oplus a}$, because there are a ways to pick an increasing sequence of length $(a-1)$on a elements.

Lemma 220

There is a map $\psi : M \to \operatorname{Hom}_R(\Lambda^a M, \Lambda^{a+1} M)$ sending m to the map $\psi(m)$, where

$$\psi(m)(m_1 \wedge \cdots \wedge m_a) = m_1 \wedge \cdots \wedge m_a \wedge m.$$

(We could have also wedged m in at the beginning.) It's easy to see that this is indeed a valid morphism and that ψ is R-linear.

Lemma 221

The map ψ described above is functorial: if $f : M \to N$ is a module morphism, then we can construct a map $\Lambda^a M \to \Lambda^{a+1} N$ in two ways which are equivalent:

$$\Lambda^{a+1} f \circ \psi(m) = \psi(f(m)) \circ \Lambda^a f.$$

Proposition 222

Let $f \in \operatorname{End}_R(R^{\oplus a})$, which is equivalent to specifying an $a \times a$ matrix with entries in R. Then $\Lambda^a f : \Lambda^a(R^{\oplus a}) \to \Lambda^a(R^{\oplus a})$ is a morphism between rank-1 modules, so taking a basis element e, we know that e must be sent to some multiple of e. We write that e is sent to $\det(f)e$, where det is the **determinant** of f.

This determinant turns out to be the usual one from linear algebra: we can check that $\det(f \circ g) = \det(f)\det(g)$, and we can check that we recover the usual formula if we let $e = e_{(1, \cdots, a)}$ and represent $f e_i = \sum_j b_{ij} e_j$ for some $b_{ij} \in R$. We will indeed find (plugging in $e_1 \wedge \cdots \wedge e_a$) that

$$\det(f) = \sum_{\sigma \in S_a} \operatorname{sgn}(\sigma) b_{1,\sigma(1)} \cdots b_{a,\sigma(a)}.$$

But continuing to fix an endomorphism f, recall also that we can send $R^{\oplus a}$ to $\operatorname{Hom}(\Lambda^{a-1} R^{\otimes a}, \Lambda^a(R^{\oplus a}))$ by Lemma 220, and in this case because we have free modules this is actually an isomorphism. We can then map $\operatorname{Hom}(\Lambda^{a-1} R^{\otimes a}, \Lambda^a(R^{\oplus a}))$ to itself by precomposing by $\Lambda^{a-1} f$. We end up getting the following commutative diagram to construct the blue map (by applying ψ, then $- \circ \Lambda^{a-1} f$, then the inverse of ψ), which we call the **adjugate** of f:

$$\begin{array}{ccc} R^{\oplus a} & \xrightarrow{\psi} & \mathrm{Hom}(\Lambda^{a-1}R^{\oplus a}, \Lambda^a R^{\oplus a}) \\ \downarrow{\scriptstyle \mathrm{adj}(f)} & & \downarrow{\scriptstyle -\circ\Lambda^{a-1}f} \\ R^{\oplus a} & \xrightarrow{\psi} & \mathrm{Hom}(\Lambda^{a-1}R^{\oplus a}, \Lambda^a R^{\oplus a}) \end{array}$$

It turns out that if we unravel the definitions,

$$\mathrm{adj}(f) \circ f = \det(f)\mathrm{id}.$$

(And with this, we can recover the matrix cofactor expression for the inverse of a matrix.) The identity is also true if we reverse the two terms on the left-hand side, but it somehow becomes more difficult to prove.

Definition 223

If $f \in \mathrm{End}_R(R^{\oplus a})$, then we can think of f as an element of $\mathrm{End}_{R[T]}(R[T]^{\oplus a})$, since $R[T]^{\oplus a}$ is really $R[T] \otimes_R R^{\oplus a}$ and we can think of f as $1 \otimes f$. If T is the multiplication by T endomorphism, then

$$\det\left(T\mathrm{id}_{R[T]^{\oplus a}} - 1 \otimes f\right)$$

is an element of $R[T]$, which we call the **characteristic polynomial** of f, which we denote $C_f(T)$.

Proposition 224 (Cayley-Hamilton)

We have $C_f(f) = 0$. (In linear algebra terms, if we plug in a matrix into its own characteristic polynomial, we will always get zero.)

Proof. By our definition above, we know that (an equality of endomorphisms)

$$\det(T\mathrm{id} - f) = \mathrm{adj}(T\mathrm{id} - f) \cdot (T\mathrm{id} - f).$$

The left-hand side is a polynomial in T with coefficients c_i in R. Meanwhile, the adjugate is an endomorphism of $R[T]^{\oplus a}$, but we can write it as a polynomial in T with coefficients B_i in $\mathrm{End}_R(R^{\oplus a})$. Setting T-coefficients equal, we see that $B_{i-1} - B_i \circ f = c_i$ for all i. But now

$$C_f(f) = \sum_i (B_{i-1} - B_i \circ f) f^i,$$

and we see that the coefficient of each f^i cancels out by telescoping sum, so $C_f(f) = 0$ as desired. □

Corollary 225 (Nakayama's lemma)

Let M be a finitely generated R-module and I be an ideal of R. Suppose that $IM = M$ (recall IM is the submodule of M in which everything is multipled by an element of I). Then there is some $r \in I$ such that $(1 + r)M = 0$.

Proof. Since M is finitely generated by some $m_1, \cdots, m_a$, we get a map $\pi : R^{\oplus a} \to M$. Then $IM = M$ means that

$$m_i = \sum a_{ij} m_j$$

for some $a_{ij} \in I$, so the linear map from the matrix $A = (a_{ij})$ is such that we actually have $\pi = \pi \circ (a_{ij})$ (as in the diagram below):

$$\begin{array}{ccc} R^{\oplus a} & \xrightarrow{\pi} & M \\ \downarrow{\scriptstyle A=(a_{ij})} & & \downarrow{\scriptstyle =} \\ R^{\oplus a} & \xrightarrow{\pi} & M \end{array}$$

So for any polynomial $f \in R[T]$, we also get a diagram commuting as below:

$$\begin{array}{ccc} R^{\oplus a} & \xrightarrow{\pi} & M \\ \downarrow{\scriptstyle f(A)} & & \downarrow{\scriptstyle f(1)} \\ R^{\oplus a} & \xrightarrow{\pi} & M \end{array}$$

In particular, taking f to be the characteristic polynomial of A and using Cayley-Hamilton, we see that $\text{char}_A(1) = 0$, so $\text{char}_A(1)M = 0$. But expanding out the polynomial, we see that $\text{char}_A(T)$ has all T-coefficients in I except the leading coefficient $1T^a$, so $\text{char}_A(1)$ is indeed $1 + r$ for some $r \in I$, as desired. □

Corollary 226

Let M be a **torsion-free**, finitely-generated R-module (meaning that for any nonzero $r \in R$ and $m \neq 0$ in M, $rm \neq 0$). Then if I is a proper ideal and $IM = M$, then $M = (0)$.

Proof. By Nakayama's lemma, we know that $(1 + r)M = (0)$ for some $r \in I$, so $(1 + r)m = 0$ for any $m \in M$. Then $1 + r \notin I$ (otherwise 1 would be in I), so in particular it is nonzero. Since M is torsion-free this means m must be zero. □

Corollary 227

Let M be a finitely-generated R-module, and suppose I is contained in all maximal ideals of R. If $IM = M$, then $M = (0)$.

Proof. We know that $r \in I$ is in all maximal ideals, but $1 + r$ is in no maximal ideal (otherwise $(1 + r) - r = 1$ would be in a maximal ideal). Thus it must be a unit, so if $(1 + r)M = (0)$ then M must indeed be (0). □

Corollary 228

Suppose M is finitely generated over R, I is an ideal contained in all maximal ideals, and $m_1, \cdots, m_a \in M$ with M generated as $\langle m_1, \cdots, m_a, IM\rangle$. Then we can suppress IM as a generator, and we in fact have $M = \langle m_1, \cdots, m_a\rangle$.

In other words, we can produce generators mod IM and use that to generate M.

Proof. Apply the previous result to the quotient $M/\langle m_1, \cdots, m_a\rangle$. □

19 November 7, 2022

We'll be discussing finitely generated modules over a PID this week, starting with the main result:

Theorem 229

Let R be a principal ideal domain, and suppose N is a submodule of a free module $R^{\oplus n}$. Then N is free, and there is a basis $e_1, \cdots, e_n$ of $R^{\oplus n}$ and elements $a_1|a_2|\cdots|a_m$ of R (with $a_m \neq 0$) such that $a_1e_1, \cdots, a_me_m$ form a basis for N. Also, these a_is are unique up to associates.

In other words, we can choose a basis for the larger module so that a subset of those basis elements, scaled appropriately, gives us a basis for the smaller module. Before proving this (which is the least important part), we'll state a few of the theorem's useful applications and work through some examples.

Corollary 230

Let R be a principal ideal domain, and let M be a finitely generated R-module. Then there are elements $a_1|a_2|\cdots|a_m$ of R (with $a_m \neq 0$), with a_1 not a unit, and some integer $d \geq 0$, such that

$$M \cong R^{\oplus d} \oplus \bigoplus_{i=1}^{m} R/(a_i).$$

Also, d, m, and (a_i) are uniquely determined up to associates by the module M. We call the a_is **invariant factors** of M.

In other words, we have some number of copies of the whole ring R. This follows from the theorem before, because we can pick some n generators for M and get a surjection $\pi : R^{\oplus n} \to \pi M$. Then applying the kernel of π to Theorem 229, we see that

$$M \cong R^{\oplus n}/\ker\pi \cong R^{\oplus(n-m)} \oplus \bigoplus_{i=1}^{m} R/(a_i),$$

and where we can drop any terms with a_i a unit (and absorb them into the free part). Notice also that the Chinese remainder theorem says that whenever $(a, b) = R$, $R/(ab) \cong R/(a) \oplus R/(b)$, so we can split up each a_i into irreducibles $\prod_\pi \pi^{m_i(\pi)}$ (times a unit). This gives us the following reformulation:

Corollary 231

Again let R be a PID and M a finitely generated R-module. Then we have

$$M \cong R^{\oplus d} \oplus \bigoplus_{(\pi) \text{ prime ideal} \neq (0)} \bigoplus_{i=1}^{n(\pi)} R/(\pi^{m_i(\pi)}),$$

where $n(\pi) = 0$ for all but finitely many π and the integers $m_i(\pi)$ are positive and in nondecreasing order. Also, M uniquely determines d, $n(\pi)$, and $m_i(\pi)$.

We'll start by thinking about the case of finitely-generated $\mathbb{Z}$-modules (so finitely generated abelian groups):

Example 232

We can list all isomorphism classes of abelian groups of order 16.

Listing out ways to split up 16 into powers of 2, we have the abelian groups

$$\mathbb{Z}/(16), \mathbb{Z}/(2) \times \mathbb{Z}/(8), \mathbb{Z}/(4) \times \mathbb{Z}/(4), \mathbb{Z}/(2) \times \mathbb{Z}/(2) \times \mathbb{Z}/(2) \times \mathbb{Z}/(2), \mathbb{Z}/(2) \times \mathbb{Z}/(2) \times \mathbb{Z}/(4).$$

(since we need a product of positive integers $a_1a_2a_3\cdots$ multiplying to 16 with none equal to 1 and $a_1|a_2|a_3|\cdots$).

Example 233

Next, we can calculate all abelian groups M that fit into a short exact sequence

$$0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to M \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0.$$

We know that M must have the form we've been discussing – we can determine some of the invariants with some algebraic manipulation. Notice that if $M \cong R^{\oplus d} \oplus \bigoplus_{i=1}^{m} R/(a_i)$, then we can localize at 0 and get $M_{(0)} \cong R_{(0)}^{\oplus d}$, so localizing our short exact sequence yields

$$0 \to \mathbb{Q} \to M_{(0)} \to \mathbb{Q} \to 0,$$

which means $M_{(0)} = \mathbb{Q}^{\oplus 2}$ just by checking dimensions and thus M must be $\mathbb{Z}^2 \oplus \bigoplus_{i=1}^{m} \mathbb{Z}/(a_i)$ (with $1 < a_1 | a_2 | \cdots | a_m \neq 0$). But now we can consider the **torsion submodule**

$$M^{\text{tor}} = \{m \in M : \exists a \neq 0 \in R \text{ with } am = 0\}.$$

The sum of two elements of M^{tor} is still in M^{tor}, because $a_1 m_1 = 0$ and $a_2 m_2 = 0$ implies $(a_1 a_2)(m_1 + m_2) = 0$. And this is closed under scalar multiplication as well. So now because R is an integral domain, $R^{\oplus d}$ has no torsion, and thus $M^{\text{tor}} \cong \bigoplus_{i=1}^{m} R/(a_i)$ – this is a way for us to "get rid of the free part." So the torsion of $\mathbb{Z} \oplus \mathbb{Z}/(5)$ is $\mathbb{Z}/(5)$, and things with torsion must go to things with torsion (because multiplying by the corresponding element of R would still kill our element even after being mapped) – this means we have an exact sequence

$$0 \to \mathbb{Z}/(5) \to M^{\text{tor}} \to \mathbb{Z}/(10)$$

(we have injectivity on the left, but we do not need to have surjectivity – consider $0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/(2) \to 0$; taking torsion gives $0 \to 0 \to 0 \to \mathbb{Z}/(2) \to 0$), but this still tells us that the order of M^{tor} must be divisible by 5 and must be a factor of 50. So that tells us that $5 | a_1 \cdots a_m | 50$ and $a_1 | \cdots | a_m$; this only works if we have one of the sequences $(a_i) = (5), (10), (25), (50), (5, 5), (5, 10)$.

So there are at most six possibilities for fitting into the short exact sequence, but we don't actually know that any of those work yet:

- Suppose we want to construct a short exact sequence $0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0$. Then we could send (a, b) to $(10a, 0, b)$ (so that the cokernel matches up with $\mathbb{Z} \oplus \mathbb{Z}/(10)$), and then we could send $(x, y, z) \to (y, x \bmod 10)$. So **this possibility can arise**.
- Next, we try $0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/(10) \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0$. This time, we can send (a, b) to $(5a, 0, 2b)$, so that the cokernel is now $\mathbb{Z}/(5) \oplus \mathbb{Z} \oplus \mathbb{Z}/(2) = \mathbb{Z} \oplus \mathbb{Z}/(10)$ (by the Chinese remainder theorem). So **this works as well** if we send (x, y, z) to $(y, 5z + 2x)$ in the second map (we can check exactness at all points explicitly).
- Now we can try $0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/(5) \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0$ – this time, we send (a, b) to $(2a, 0, b, 0)$, so the cokernel is $\mathbb{Z}/(2) \oplus \mathbb{Z} \oplus (0) \oplus \mathbb{Z}/(5)$, which is again correct. Then we send (x, y, z, w) to $(y, 2w + 5x)$ much like before, and **this is also possible**.
- Similarly, we can construct $0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/(25) \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0$ by sending (a, b) to $(2a, 0, 5b)$ – the cokernel works out again, and we need to send (x, y, z) to $(y, 5x + 2(z \bmod 5))$, giving us another **valid exact sequence**.
- It is easy to construct $0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/(5) \oplus \mathbb{Z}/(10) \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0$ by sending (a, b) to $(a, 0, b, 0)$ and then sending (x, y, z, w) to (y, w) – this is clearly **exact** too.

- Finally, we can construct $0 \to \mathbb{Z} \oplus \mathbb{Z}/(5) \to \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/(50) \to \mathbb{Z} \oplus \mathbb{Z}/(10) \to 0$ by sending (a, b) to $(a, 0, 10b)$, giving us yet another **exact sequence**.

Basically, we can just try to construct two valid maps by checking that cokernel is of the right type, show injectivity and surjectivity of the first and second maps, and then check exactness at the middle. So in this case, this strategy works, but it's generally surprising how many of the possibilities will work and there typically has to be a "good reason" for it to fail. (For example, one situation where we do run into issues is that M cannot use more generators than the first and last modules combined.)

Our next application will be to linear algebra:

Example 234

Let K be a field, and let V be a finite-dimensional K-vector space. Given a linear map $T \in \mathrm{End}_K(V)$, we can think of V as a $K[x]$-module in which x acts by T, meaning that $\left(\sum a_i x^i\right) v = \sum a_i T^i(v)$. Then V is finitely generated over $K[x]$ (because it was already finitely generated over K), so we have an isomorphism as $K[x]$-modules

$$V \cong K[x]^{\oplus d} \oplus \bigoplus_{i=1}^{m} K[x]/(a_i)$$

with $a_1 | \cdots | a_m \neq 0$ are all polynomials over K, which we can assume to be monic by appropriately multiplying, and a_1 is not a unit (meaning it has positive degree).

In such a case, we know that d is actually zero, because $K[x]$ is already infinite-dimensional over K and V has to be finite-dimensional, so we just have a direct sum of $K[x]/(a_i)$s. Furthermore, $K[x]/(a_i)$ has dimension $\deg(a_i)$ (it has a K-basis $\{1, x, \cdots, x^{\deg(a_i)-1}\}$ by the division algorithm), so we must have $\dim V = \sum_{i=1}^{m} \deg(a_i)$. And since T acts as multiplication by x, the action of T on each $K[x]/(a_i)$ is also multiplication by x.

Now if we have another vector space V' with action T', we can write $V' \cong \bigoplus_{i=1}^{m'} K[x]/(a_i)$. Then there is an isomorphism $f : V \xrightarrow{\sim} V'$ if and only if $f \circ T = T' \circ f$ (in other words, there is a linear map that commutes with the structure of $K[x]$-modules), and we know this only happens if $m = m'$ and $a_i = a_i'$ for all i. So this gives us a way to test whether two vector spaces with an endomorphism are isomorphic with that endomorphism. We'll talk more about this next time!

20 November 9, 2022

Last time, we stated the classification theorem for finitely generated modules M over a PID R, saying that such modules always take the form $M \cong R^{\oplus d} \oplus \bigoplus_{i=1}^{m} R/(a_i)$, where d is some nonnegative integer, $a_1 | \cdots | \cdots a_m$ (with a_1 not a unit and a_m nonzero), where the ideals (a_i) are uniquely determined (meaning the a_is are determined up to associates) and so are d and m. We saw some examples in the context of abelian groups, and we started looking at the case where $R = K[x]$ (in which case giving a module is like giving a vector space over K plus an endomorphism on that space). In particular, where V is finite-dimensional over K, we know there are $a_1 | \cdots | a_m$ in $K[x]$ (we called these **invariant factors** and we can take them to be monic) with all degrees positive, such that

$$V \cong \bigoplus_{i=1}^{m} K[x]/(a_i(x)),$$

where the action of our endomorphism T is multiplication by x (in each direct summand) and $\dim V = \sum_{i=1}^{m} \dim K[x]/(a_i(x)) = \sum_{i=1}^{m} \deg(a_i)$. This then lets us write down a matrix form for T (multiplying by x): it is block diagonal (with one block

for each a_i), where if we pick the monomials as a basis, each block looks like

$$T\Big|_{K[x]/a_i(x)} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_{i,0} \\ 1 & 0 & \cdots & 0 & -a_{i,1} \\ 0 & 1 & \cdots & 0 & -a_{i,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{i,\deg(a_i)-1} \end{bmatrix}.$$

(The last row here comes from the fact that $a_{i,0} + a_{i,1}x + \cdots + a_{i,\deg(a_i)-1}x^{\deg(a_i-1)} + x^{\deg(a_i)} = 0$.) A matrix with diagonal blocks that look like this is said to be in **rational canonical form**, and it is unique because given a matrix of this form representing T (where $a_1(x)|a_2(x)|\cdots|a_m(x)$ we know that each block is isomorphic to $K[x]/(a_i(x))$ and the conditions are satisfied. And from this we can also calculate the characteristic polynomial of T – it turns out that we actually have

$$\mathrm{char}_T(x) = a_1(x)\cdots a_m(x).$$

Indeed, we must calculate the polynomials

$$\det \begin{bmatrix} x & 0 & \cdots & 0 & -a_{i,0} \\ 1 & x & \cdots & 0 & -a_{i,1} \\ 0 & 1 & \cdots & 0 & -a_{i,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x - a_{i,\deg(a_i)-1} \end{bmatrix} \stackrel{?}{=} a_0 + a_1x + \cdots + a_dx^d$$

and multiply them together across all i to get the result, and this is indeed true by induction if we expand the determinant along the first column.

Lemma 235

There exists a monic polynomial $m(x)$ such that $m_T(T) = 0$ and such that whenever $f(T) = 0$ we have $m_T|f$; we call this the **minimal polynomial**.

In the particular case above, because $a_1|\cdots|a_m$, the minimal polynomial is $m_T(x) = a_m(x)$.

Proof. Consider the annihilator

$$\mathrm{Ann}_{K[x]}(V) = \{f \in K[x] : f(T) = 0\};$$

since the annihilator is an ideal and R is a PID, this must be equal to (m_T) for some m_T unique up to units and thus unique if we insist that it is monic. □

Last time, we mentioned that for any (V, T) and (V', T') (for V, V' finite-dimensional K-vector spaces) and T, T' corresponding endomorphisms, we see that $V \cong V'$ as a $K[x]$ module if and only if there is some isomorphism $g : V \to V'$ with $g \circ T = T' \circ g$ (meaning that we have the same number of invariant factors of T and of T' and $a_i = a_i'$ for all i). This can in fact be restated in terms of matrices:

Definition 236

Two matrices $A, B \in M_{n\times n}(K)$ are **similar** if they are conjugate, meaning that there is some $g \in \mathrm{GL}_n(K)$ with $B = gAg^{-1}$.

In other words, similarity means that the $K[x]$-modules $(K^{\oplus n}, A)$ and $(K^{\oplus n}, B)$ are isomorphic via a map g, since that is the same as saying that $Bg = gA$. And that's then equivalent to saying that A and B must have the same invariant factors. (So in some sense, this is a better version of Jordan normal form.)

Example 237

With this method, we can determine the conjugacy classes in a group like $\mathrm{GL}_3(\mathbb{F}_2) = \mathrm{GL}_3(\mathbb{Z}/(2))$.

It suffices to look for monic invariant factors $a_1|\cdots|a_m \in \mathbb{Z}/(2)[x]$, such that all degrees are positive and $\sum \deg(a_i) = 3$. But we also need an additional condition – remember that the constant term of the characteristic polynomial is the determinant, which we want to be nonzero because we want invertible matrices. Thus we must require that x does not divide any a_m.

But now we can just list possibilities: in one case we can have $m = 1$, meaning we have a single cubic polynomial with nonzero constant term. The possibilities here are $x^3+1, x^3+x^2+1, x^3+x+1$, and x^3+x^2+x+1. Alternatively, we can have $m = 2$, in which case we must have a linear and a quadratic polynomial. So then a_1 must be $x+1$ and $a_2 = (x+1)^2 = x^2+1$. Finally, if $m = 3$, then we can have $a_1 = a_2 = a_3 = x+1$. Putting all of this together, we see that there are six conjugacy classes of matrices. We can then go back to our rational canonical form and find a representative for each class: in the order we listed them, they are

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(colors indicating diagonal blocks in the last two cases).

Example 238

Next, we'll determine the number of conjugacy classes of $g \in \mathrm{GL}_3(K)$ such that $g^5 = 1$ (just to avoid having to write out the matrices).

This time, we're again looking for monic polynomials $a_1|\cdots|a_m$ such that all degrees are positive and $\sum \deg(a_i) = 3$, but if $g^5 = 1$ that means the minimal polynomial $m_g(x) = a_m(x)$ must divide $x^5 - 1$ (and in particular this does mean the constant terms will be nonzero so we will get an invertible matrix).

We'll first do the case where $K = \mathbb{C}$. Factoring over the complex numbers,

$$x^5 - 1 = (x-1)(x-\zeta)(x-\zeta^2)(x-\zeta^3)(x-\zeta^4)$$

where $\zeta = e^{2\pi i/5}$. We can again do casework on m: for $m = 1$, we take any cubic factor of $x^5 - 1$, so there are $\binom{5}{3} = 10$ ways to pick three distinct ζ^is. For $m = 2$, we pick any linear factor (in 5 ways), and then we pick any quadratic factor which includes that linear factor (in 4 more ways), giving us $5 \cdot 4 = 20$ conjugacy classes. Finally, for $m = 3$ we just have 5 ways to pick a single linear factor three times. This gives us $10 + 20 + 5 = \boxed{35}$ conjugacy classes in total.

Next, we can consider the case $K = \mathbb{R}$. This time for $m = 1$, we again want $a_1(x) + (x-\zeta_1)(x-\zeta_2)(x-\zeta_3)$, but in order for the polynomial to be real all roots must come in conjugate pairs. This means we must have 1 and a pair of the other roots, giving us 2 possibilities. But for $m = 2$, there are no possibilities – we would need to have $a_1(x) = (x-1)$, but then $a_2(x)$ is $(x-1)(x-\zeta^i)$ for some $\zeta^i \neq 1$ but that will never be real. And finally for $m = 3$ there is only one possibility where all factors are $(x-1)$. Thus there are only $2 + 1 = \boxed{3}$ conjugacy classes in this case.

After that, we consider $K = \mathbb{F}_2$. This time, the polynomial can be factored as

$$x^5 - 1 = (x-1)(x^4 + x^3 + x^2 + x + 1),$$

and this time $x^4 + x^3 + x^2 + x + 1$ turns out to be irreducible (because there are no linear factors, and we can check that no product of quadratic factors works either – the only possibility would have been $(x^2 + x + 1)^2$ to avoid roots, and that doesn't work). So now $m = 1$ and $m = 2$ are not possible (because we can't have a quadratic or cubic factor), and the only possibility is $(x-1)$ three times and there is only $\boxed{1}$ conjugacy class.

Finally, consider $K = \mathbb{F}_5$, in which case $x^5 - 1 = (x-1)^5$. And in this case each of $m = 1, 2, 3$ has a unique possibility ($(x-1)^3$ for $m = 1$, $(x-1), (x-1)^2$ for $m = 2$, and $(x-1), (x-1), (x-1)$ for $m = 3$), so we have $\boxed{3}$ conjugacy classes.

Remark 239. *Remember that we have been calculating conjugacy classes of matrices satisfying $g^5 = 1$, but when we consider conjugacy we can still conjugate by any matrix in $GL_3(K)$. (Indeed, the set of matrices where $g^5 = 1$ doesn't form a group because $GL_3(K)$ is nonabelian.)*

Next time, we'll briefly consider Jordan normal form and then sketch the proof of this clasification theorem.

21 November 11, 2022

We'll start with Jordan normal form today – start with any algebraically closed field K. Then the only irreducible polynomials in $K[x]$ are the linear ones $X - \lambda$ (since any polynomial has a root and we can pull out the corresponding linear factors). Thus if V is a finite-dimensional K-vector space, and T is a K-linear endomorphism of V (recall this is the same thing as having a $K[x]$-module V), then we have a normal form for finitely generated $K[x]$-modules in two ways. One is where we choose polynomials $a_1 | \cdots | a_n$, and the other is where we choose irreducible polynomials π and get

$$R^{\oplus d} \oplus \bigoplus_{(\pi)} \bigoplus_{i} R/(\pi^{m_{i,\pi}}).$$

We'll use this latter description here, and we then find that over an algegbraically closed K, we must have

$$V \cong \bigoplus_{\lambda \in K} \bigoplus_{i=1}^{n_\lambda} K[x]/(x-\lambda)^{m_{\lambda,i}},$$

where $m_{\lambda,i}$ are positive and $n_\lambda = 0$ for all but finitely many of the λs. Then we can also describe the action of T quite easily – if we choose the basis $\{1, (x-\lambda), \cdots, (x-\lambda)^{m_\lambda - 1}\}$ for each of these summands, then notice that

$$x(x-\lambda)^j = (x-\lambda)^{j+1} + \lambda(x-\lambda)^j,$$

so we can choose a basis so that our matrix for T will be block diagonal (blocks corresponding to the different direct summands) of the form

$$\begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 1 & \lambda \end{bmatrix}.$$

So any matrix in $M_{n\times n}(K)$ (thought of as an endomorphism acting on $K^{\oplus n}$) will be similar (conjugate) to a block diagonal matrix with such **Jordan blocks**, with the λs, n_λs, and $m_{\lambda,i}$s uniquely determined. So we have uniqueness up

to reordering the blocks, and we call this representation **Jordan normal form**. It has the disadvantage of only working over algebraically closed field, but it's more useful for computing matrix powers than rational canonical form.

We're now ready to turn back to the main classification theorem we stated two lectures ago and do the proof. Recall the statement: for any PID R, any submodule N of a free module $R^{\oplus n}$ is free, and we can choose a basis $y_1, \cdots, y_n$ of $R^{\oplus n}$ and $a_1|a_2|\cdots|a_m$ elements of R (with $a_m \neq 0$) such that $a_1e_1, \cdots, a_me_m$ form a basis for N.

Proof. We can assume N is nonzero (otherwise there's nothing to prove). If this result were true, then for any linear map $R^{\oplus n} \to R$, everything in N would get sent to a multiple of a_1 (since $a_1|a_2|\cdots$), so we want to identify a_1 as the largest element of R with this property. Towards that, define

$$\mathcal{X} = \{\phi N \lhd R : \phi : R^{\oplus n} \to R \text{ is } R\text{-linear}\}.$$

(Unpacking the notation, each element of this set is the image of N under an R-linear map to R, so it's an ideal of R.) This contains a nonzero ideal because we can look at the projections π_i onto the ith coordinate; those cannot all be zero if N is nonzero. Since R is a principal ideal domain, it is noetherian, and thus we can pick some maximal element $\phi_1 N = (a_1)$. In particular, this means there is some $y \in N$ such that $\phi_1(y) = a_1$ (we expect it to be a_1y_1).

So now take $\psi : R^{\oplus n} \to R$ to be any other linear map. If we look at the ideal $(a_1, \psi(y))$ in R, because R is a PID it must be (d), and we must be able to write $x = \alpha a_1 + \beta\psi(y) = (\alpha\phi_1 + \beta\psi)(y)$. So $(\alpha\phi_1 + \beta\psi)(N)$ contains (d), which contains (a_1), but now $(\alpha\phi + \beta\psi)(N)$ is also an element of $\mathcal{X}$ so by maximality all of these ideals must be equal. So that means $(a_1) = (d)$ and a_1 must divide $\psi(y)$, as desired. In particular, choosing ψ to be each of the coordinate projections in turn, we see that a_1 must divide every coordinate of y. So we can indeed write $y = a_1y_1$ for some $y_1 \in R^{\oplus n}$, which gives us our first basis element. (And we have $\phi_1(y_1) = 1$, so in some sense we can't divide y_1 any further.)

So now if $m \in R^{\oplus n}$, we can write m as a linear combination of y_1 and the rest, which we denote $m = \phi_1(m)y_1 + (m - \phi_1(m)y_1)$. But the latter term is in the kernel of ϕ_1 (because $\phi_1(y_1) = 1$), so $R^{\oplus n} = Ry_1 + \ker(\phi_1)$. And this is in fact a direct sum because being in both Ry_1 and $\ker(\phi_1)$ would make the element zero. Thus $\boxed{R^{\oplus n} = Ry_1 \oplus \ker(\phi_1)}$, and we want to describe how N looks under this decomposition too. For any element $m \in N$, we know $\phi_1(m)$ is divisible by a_1, so $\boxed{N = Ra_1y_1 \oplus (N \cap \ker(\phi_1))}$. Thus we've actually managed to split up N in the same way as $R^{\oplus n}$, which is the goal.

Unfortunately we can't just apply induction directly from here, since we don't necessarily know that $\ker(\phi_1)$ (the bigger module) is free yet. So we have to proceed in two steps here: first, we show that **any** submodule N (not just the specific N in the statement) is free by induction on the dimension of the localization $N_{(0)}$ over the quotient ring $Q(R)$. For the base case, if $N_{(0)} = (0)$ then $N = (0)$. Then for the inductive step, we know that $N = Ra_1y_1 \oplus (N \cap \ker(\phi))$ from above, so localizing at zero yields

$$N_{(0)} = Q(R) \oplus (N \cap \ker(\phi))_{(0)}.$$

The dimension of $(N \cap \ker(\phi))_{(0)}$ is one less than that of $N_{(0)}$, so (the unlocalized) $(N \cap \ker(\phi))$ must be free by the inductive hypothesis. So N is the direct sum of two free modules and thus N itself is free.

So now we can complete the proof: we induct on n. If $n = 0$ there is nothing to prove, and in general we use the boxed decomposition above. We have just shown that $\ker(\phi_1)$ is free (because it is a submodule of the bigger module so the argument above holds), and rank is additive so $\ker\phi_1$ has rank $(n-1)$. Thus there is a basis $y_2, \cdots, y_n$ of $\ker(\phi_1)$ and $a_2|\cdots|a_m \neq 0$ so that $N \cap \ker(\phi_1)$ has basis $a_2y_2, \cdots, a_my_m$. From the first boxed statement we know that $y_1, y_2, \cdots, y_m$ do form a basis of $R^{\oplus n}$, and from the second we see that $a_1y_1, \cdots, a_my_m$ is a basis of N. So we just need to show that a_1 divides a_2; indeed, take the map ψ which sends y_1 and y_2 to 1 and y_i to 0 for any other i.

Then $\psi(N) = (a_1, a_2)$ is an element of $\mathcal{X}$ containing (a_1), but by maximality these must be equal and thus $a_1|a_2$. $\square$

And as we mentioned in a previous lecture, this allows us to recover that for any finitely generated R-module M over a PID R, we do have

$$M \cong R^{\oplus d} \oplus \bigoplus_{i=1}^{m} R/(a_i)$$

by applying our classification theorem to the kernel of the map $R^{\oplus n} \to M$ (and noting that we can throw away any of the terms here where a_i is a unit, since $R/(a_i) = 0$). The remaining thing to show is that these $d, m, (a_i)$ are **uniquely** determined by M:

Lemma 240

If $c_1|c_2|\cdots|c_t$, where c_1 is not a unit in R, then t is the minimal number of generators for $M = R/(c_1)\oplus\cdots\oplus R/(c_t)$.

Proof. Choose a maximal ideal $\mathfrak{m}$ containing (c_1) (the latter is a proper ideal because c_1 is not a unit). If M can be generated by s elements, then so can $M/\mathfrak{m}M \cong R/\mathfrak{m} \oplus \cdots \oplus R/\mathfrak{m}$. But this is now a vector space over a field, and a vector space of dimension t can only be generated if we have at least t elements, so $s \geq t$. $\square$

And we can now finally show uniqueness:

Proposition 241

Let R be a PID. If M can be written as $R^{\oplus d} \oplus \bigoplus_{i=1}^{m} R/(a_i) \cong R^{\oplus e} \oplus \bigoplus_{i=1}^{n} R/(b_i)$, and $a_1|\cdots|a_m$ and $b_1|\cdots|b_n$ with a_1, b_1 not units and $a_m, b_n \neq 0$, then $d = e$, $m = n$, and $(a_i) = (b_i)$ for all i.

Proof. Localizing at zero, we have $Q(R)^{\oplus d} \cong Q(R)^{\oplus e}$, so $d = e$ by invariance of dimension of a vector space. Now M has a torsion submodule, which is $\bigoplus R/(a_i) \cong \bigoplus R/(b_i)$, so we can say without loss of generality that $d = e = 0$.

But now $\bigoplus_{i=1}^{m} R/(a_i)$ is generated by m things, and $\bigoplus_{i=1}^{n} R/(b_i)$ is generated by n things, so by our lemma above $m \geq n$ and $n \geq m$, so $n = m$. Now for any element $a \in R$, define the ***a*-torsion**

$$M[a] = \{m \in M : am = 0\}.$$

Looking at $M/M[a]$, we see that

$$(R/(a_i))[a] = \{r + (a_i) : ar \in (a_i)\} = \{r + (a_i) : a_i|ar\},$$

which is the same thing as requiring that $\frac{a_1}{\gcd(a_i,a)}|\frac{a}{\gcd(a_i,a)}r$, meaning that we require $\frac{a_i}{\gcd(a_i,a)}|r$. Thus $(R/(a_i))[a] = (a_i/(\gcd(a_i, a)))/(a_i)$, so

$$\bigoplus R/((b_i/\gcd(b_i, a))) \cong M/M[a] \cong \bigoplus R/((a_i/\gcd(a_i, a))).$$

Then the minimum number of generators on the left is the number of b_is that don't divide a, and the minimum number of generators on the right is the number of a_is that don't divide a. But now taking $a = a_1$ (and using that $m = n$), we see that all b_is must divide a_1 and in particular $b_1|a_1$. Similarly $a_1|b_1$, so a_1 and b_1 are associates. More generally taking $a = a_i$ allows us to show that a_i and b_i are associates for all i, so $(a_i) = (b_i)$. $\square$

22 November 14, 2022

In these last three weeks, we'll briefly introduce **homological algebra** – this is somewhat difficult to motivate but turns out to be very useful in various branches of mathematics (including algebra, algebraic geometry, and algebraic topology). We'll decide a setting in which we'll do homological algebra, and the most useful one is to work in **abelian categories**. Recall that the category **R-mod** has a few special properties, namely that we can add morphisms and describe kernels and cokernels, that finite products and finite coproducts are the same, and so on. These can all be abstracted into a more general concept – for the sake of presentation, we'll **state results in abelian categories and prove them in R-mod**.

Definition 242

A category $\mathcal{C}$ is **additive** if it satisfies the following properties:

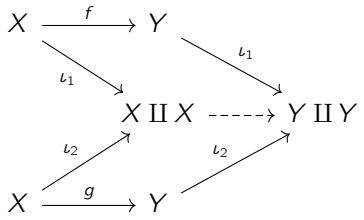
- $\mathcal{C}$ has an initial object (an object that maps uniquely to any object) and a final object (an object for which any object maps to it uniquely), and the map from the initial object to the final object is **required to be an isomorphism**. We then call this (initial/final) object the **null object** and denote it (0). In particular, composing the maps $X \to (0) \to Y$, we get a unique **zero morphism** $0 : X \to Y$ for any $X, Y \in \mathcal{C}$.
- Binary products and coproducts exist, meaning that we have $X \times Y$ along with its projection maps p_x, p_y to X and Y, as well as $X \amalg Y$ along with maps from X and Y into it. So if we take the identity map $\mathrm{id}_X : X \to X$ and the zero map $0 : X \to Y$, we get a unique map $j_X : X \to X \times Y$ (this is intuitively "embedding into the first factor and zero in the second") and similarly $j_y : Y \to X \times Y$. Thus by the universal property of the coproduct we get a map $\alpha : X \amalg Y \to X \times Y$. We then also **require α to be an isomorphism** – we then call this object the **direct sum** $X \otimes Y$.

So in an additive category, thinking of $X \oplus X$ as a product, we have two natural maps $p_1 : X \oplus X \to X$ and $p_2 : X \oplus X \to X$ (the projections onto the two coordinates). Then taking the identity maps from $X \to X$ in each coordinate, we get a unique map $\Delta : X \to X \oplus X$ commuting with those maps, which we call the **diagonal map**. Similarly, thinking of $X \oplus X$, the identity maps from the two X's into X gives us a unique map $+ : X \oplus X \to X$ – this can be thought of as the addition map.

If we now have two maps $f, g : X \to Y$, we can consider the composite map

$$X \xrightarrow{\Delta} X \oplus X \xrightarrow{(f,g)} Y \oplus Y \xrightarrow{+} Y.$$

(To explain the middle map, consider the following commutative diagram:



We get two maps from X to $Y \amalg Y$, which induces a map $X \amalg X \to Y \amalg Y$.) So now we get a new map $f + g : X \to Y$, and it turns out that $(+, 0)$ make $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ into an abelian group for any $X, Y \in \mathcal{C}$. Furthermore, $\mathrm{Hom}_{\mathcal{C}}(X, Y) \times \mathrm{Hom}_{\mathcal{C}}(Y, Z) \to \mathrm{Hom}_{\mathcal{C}}(X, Z)$, defined via composition of maps, will be bilinear.

Remark 243. *Some additional conditions are required for this part – in general what we've written here isn't enough to guarantee additive inverses – and this is clarified in a few lectures.*

Definition 244

We call a functor $F : \mathcal{C} \to \mathcal{D}$ between additive categories **additive** if it preserves finite products and coproducts (including the initial and final object, which are the empty coproduct and product). In particular, this means $F : \mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(FX, FY)$ is a group homomorphism.

This way of setting up abelian categories is nice because it shows that we don't need to additionally define addition as a new structure – it's just following from the axioms. But there is an alternative way to describe all of this: we can say that a **pre-additive category** is a category in which each $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ is endowed with the structure of an abelian group, such that the compositions $\mathrm{Hom}_{\mathcal{C}}(X, Y) \times \mathrm{Hom}_{\mathcal{C}}(Y, Z) \to \mathrm{Hom}_{\mathcal{C}}(X, Z)$ are all bilinear. Then in a pre-additive category, a functor $F : \mathcal{C} \to \mathcal{D}$ is called additive if $F(f + g) = F(f) + F(g)$ for all $f, g : X \to Y$ for $X, Y \in \mathcal{C}$.

Proposition 245

A category $\mathcal{C}$ is additive if and only if it is pre-additive and has finite products and coproducts.

Proof. The forward direction is clear. For the reverse direction, notice that we have an isomorphism between the initial and final objects, because the composite map initial $\to$ final $\xrightarrow{0}$ initial is the identity (since we always have a zero map from the final object to the initial object because any abelian group has the zero element) and similarly final $\xrightarrow{0}$ initial $\to$ final is the identity, so we do have an isomorphism initial $\to$ final with inverse map 0. And we also have to show that the map $\alpha : X \amalg Y \to X \times Y$ is an isomorphism, but looking back to how we defined α: we have $p_X \alpha \iota_X = \mathrm{id}_X, p_Y \alpha \iota_Y = \mathrm{id}_Y, p_X \alpha \iota_Y = 0, p_Y \alpha \iota_X = 0$. But we get a map $\iota_X \circ p_X + \iota_Y \circ p_Y$ (addition coming because of the abelian group structure given by being preadditive), and $(\iota_X \circ p_X + \iota_Y \circ p_Y)(\alpha)$ is the identity because we can compose those maps with ι_X and ι_Y, seeing that

$$(\iota_X \circ p_X + \iota_Y \circ p_Y)(\alpha)\iota_X = \iota_X + 0 = \iota_X,$$

$$(\iota_X \circ p_X + \iota_Y \circ p_Y)(\alpha)\iota_Y = 0 + \iota_Y = \iota_Y.$$

Thus by the universal product of the coproduct, $(\iota_X \circ p_X + \iota_Y \circ p_Y)(\alpha) = \mathrm{id}_{X \amalg X}$. A very similar analysis shows that this holds when we switch the order of terms on the left.

We should check that the additive structures coincide with our two definitions, but we won't do that here – it does turn out that they are exactly equivalent. □

Definition 246

An additive category $\mathcal{C}$ is **abelian** if it satisfies the following properties:

- If $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, then the combination of maps $f : X \to Y$ and $0 : X \to Y$ has a limit $\ker(f) \to X$ (which we call the **kernel** of f) and a colimit $Y \to \mathrm{coker}(f)$ (which we call the **cokernel** of f). We can check that these are indeed a generalization of the definitions for R-modules, the kernel is always a monomorphism, and the cokernel is always an epimorphism.
- Any monomorphism is a kernel and any epimorphism is a cokernel.

Example 247

As discussed, **R-mod** is an example of an abelian category. Somewhat similarly, if Γ is a group, then we can define Γ-**mod** to be the category of abelian groups with an action of Γ (where morphisms must commute with the action of Γ), and this is also an abelian category.

Example 248 (Not on quals syllabus but useful for motivation)

Let X be a topological space. Then **Open(X)** is the category in which objects are open sets of X and morphisms are inclusions of open sets (so there is always either zero or one morphism between two objects). A **presheaf** (of abelian groups) on X is a contravariant functor $\mathcal{F} : \mathbf{Open}(X) \to \mathbf{Ab}$.

More concretely, for every open set $U \in X$ we associate an abelian group $\mathcal{F}(U)$, such that whenever $V \subset U$ we also have a **restriction map** $\mathcal{F}(U) \to \mathcal{F}(V)$ sending m to $m|_V$, such that the restriction $\mathcal{F}(U) \to \mathcal{F}(U)$ is the identity and the restriction is compatible with triples (meaning that if $W \subset V \subset U$ are open in X, then the composition of restrictions $\mathcal{F}(U) \to \mathcal{F}(V) \to \mathcal{F}(W)$ yields the same result as the restriction $\mathcal{F}(U) \to \mathcal{F}(W)$. Then a morphism between presheaves $\mathcal{F}$ and $\mathcal{G}$ is a natural transformation $f : \mathcal{F} \to \mathcal{G}$ – in other words, for every $U \subset X$, we want a map $f_U : \mathcal{F}(U) \to \mathcal{G}(U)$ compatible with restriction (so applying f_U and then restricting to V is the same as restricting to V and then applying f_V).

Definition 249

A presheaf $\mathcal{F}$ is a **sheaf** if the following holds: for all $U \subset X$ open and any open covers $\{U_i\}_{i \in I}$ of U, then there are two maps $\prod_{i \in I} \mathcal{F}(U_i) \to \prod_{(i,j) \in I \times I} \mathcal{F}(U_i \cap U_j)$, namely (1) the one sending (s_i) to $(s_i|_{U_i \cap U_j})_{(i,j)}$ and (2) the one sending (s_i) to $(s_j|_{U_i \cap U_j})_{(i,j)}$. Then we require that the component-wise restriction $\mathcal{F}(U) \to \prod_{i \in I} \mathcal{F}(U_i)$ is isomorphic to the limit of the two maps (1) and (2).

The idea is that we can basically "determine $\mathcal{F}$" locally, and another way to say this sheaf condition is that (1) if $s \in \mathcal{F}(U)$ and $s|_{U_i} = 0$ for all i, then $s = 0$, and (2) if $s_i \in \mathcal{F}(U_i)$ for all i and these s_is (called **sections**) are compatible on intersections, meaning $s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}$, then there is some $s \in \mathcal{F}(U)$ such that $s|_{U_i} = s_i$ for all i (and in fact by (1) this must be unique if it exists).

We thus have the categories **Sh(X)** and **PreSh(X)** (the category of sheaves and category of presheaves on X) with $\mathbf{Sh}(X) \subset \mathbf{PreSh}(X)$ by inheriting all of the objects.

Example 250

Let $\mathcal{C}(U)$ be the set of all continuous functions $f : U \to \mathbb{C}$ with f continuous. This is a presheaf because we can restrict a continuous function to any open subset, and it's also a sheaf because a function is zero if and only if it is zero everywhere locally. Indeed, if $f_i : U_i \to \mathbb{C}$ are continuous and f_i, f_j agree on intersections $U_i \cap U_j$, then we have a continuous function $f : U \to \mathbb{C}$ on all of $U = \bigcup_{i \in I} U_i$ (just by defining $f(x) = f_i(x)$ if $x \in U_i$).

Example 251

The constant sheaf $\underline{\mathbb{C}}(U)$ is the set of functions $f : U \to \mathbb{C}$ that are locally constant (meaning that around every point there is a neighborhood on which f is constant) – since this doesn't depend on the topology of $\mathbb{C}$, for any abelian group we can also similarly define $\underline{A}(U) = \{f : U \to A, f \text{ locally constant}\}$.

23 November 16, 2022

Last time, we mentioned the categories of presheaves and sheaves on a topological space, which assign to each open set U an abelian group $\mathcal{F}(U)$ in a way that lets us restrict a section $s \in \mathcal{F}(U)$ to some $s|_V \in \mathcal{F}(V)$ whenever $V \subset U$ (which is compatible with repeated restriction). And the condition for being a sheaf is that this construction is "local:" if $U = \bigcup_{i \in I} U_i$ and we have $s_i \in \mathcal{F}(U_i)$ such that $s_i = s_j$ on $U_i \cap U_j$, then we have a unique $s \in \mathcal{F}(U)$ which restricts to s_i on each U_i.

Definition 252

Let $\mathcal{F}$ be a presheaf, and let $x \in X$. The **stalk** of $\mathcal{F}$ at x is the direct limit $\mathcal{F}_x = \varinjlim_{U \ni x \text{ open}} \mathcal{F}(U)$.

For example, if $\mathcal{F}$ is the set of continuous functions from U to $\mathbb{C}$, then the stalk would be the continuous functions defined on a neighborhood of x, where two functions are equal if they're equal on some small enough set (sometimes this is called the "germs of continuous functions at x" in analysis terminology). And for the constant sheaf $\underline{A}$, we have $\underline{A}_x = A$.

Fact 253

We can detect monomorphisms in sheaves in the obvious way: if $\mathcal{F} \to \mathcal{G}$ is a monomorphism, that's equivalent to having a monomorphism (for abelian groups, injection) $\mathcal{F}(U) \hookrightarrow \mathcal{G}(U)$ for all $U \subset X$ open, which is equivalent to $\mathcal{F}_x \hookrightarrow \mathcal{G}_x$ for all $x \in X$. But we have to be more careful with epimorphisms: $\mathcal{F} \to \mathcal{G}$ is an epimorphism if and only if $\mathcal{F}_x \to \mathcal{G}_x$ is surjective for all $x \in X$, but it's possible to have an epimorphism where the map $\mathcal{F}(U) \to \mathcal{G}(U)$ is not surjective.

Returning now to general abelian categories, we'll introduce some terminology similar to that which we previously defined for R-modules:

Definition 254

Let $\mathcal{C}$ be an abelian category, and suppose we have a series of morphisms between objects of $\mathcal{C}$:

$$\cdots \to X_0 \xrightarrow{f_0} X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} X_3 \to \cdots.$$

If $f_{i+1} \circ f_i$ for all i, we say that this sequence is a **complex**, and if $\operatorname{im}(f_i) = \ker(f_{i+1})$ for all i, we say this sequence is **exact**.

We didn't formally define the image in a general abelian category last time, but we know that we have a map $\ker f \to X \xrightarrow{f} Y \to \operatorname{coker}(f)$ for any morphism $f : X \to Y$. Then we can look at either $\operatorname{coker}(\ker f \to X)$ or $\ker(Y \to \operatorname{coker}(f))$, and they turn out to be equal and we call that the **image** of f.

Definition 255

An exact sequence of the form $0 \to X \to Y \to Z \to 0$ is called **short exact**. Similarly, an exact sequence $0 \to X \to Y \to Z$ is called **left exact** (and equivalent to the statement $X = \ker(Y \to Z)$), and an exact sequence $X \to Y \to Z \to 0$ is **right exact** (and equivalent to $Z = \operatorname{coker}(X \to Y)$).

Being short exact is equivalent to being both left and right exact, and a long exact sequence can always be split into a sequence of short exact sequences: if we have a long exact sequence $\cdots \to X_0 \xrightarrow{f_0} X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} X_3 \to \cdots$,

that's equivalent to requiring that we have a short exact sequence

$$0 \to \operatorname{coker}(f_{i-2}) \to X_i \to \ker(f_{i+1}) \to 0$$

for all i. (So for example, at $i = 2$ above, we need that the cokernel of $f_0 : X_0 \to X_1$, which is $X_1/\operatorname{im}(f_0) \cong X_1/\ker(f_1)$, and the kernel of f_3, which is the image of $f_2 : X_2 \to X_3$, to fit into the sequence.)

Recall that additive functors preserve the zero object and additive direct sums, but they are not required to (and may not) preserve kernels and cokernels. So homological algebra is about how additive functors interact with those kernels and cokernels:

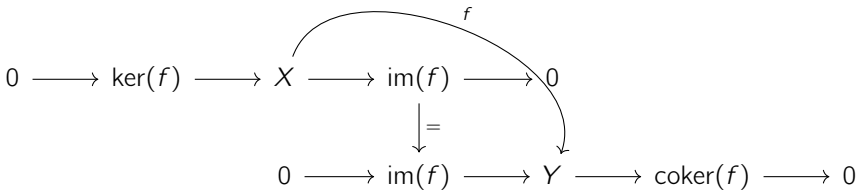
Example 256

Consider the functor **Ab** $\to$ **Ab** given by tensoring $\otimes_{\mathbb{Z}}\mathbb{Z}/(2)$. Then $0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/(2) \to 0$ is short exact, but tensoring gives us a sequence $0 \to \mathbb{Z}/(2) \xrightarrow{0} \mathbb{Z}/(2) \xrightarrow{\text{id}} \mathbb{Z}/(2) \to 0$, and the zero map $\mathbb{Z}/(2) \to \mathbb{Z}/(2)$ is not injective. So tensoring does not preserve kernels. Similarly, the functor $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/(2), -)$ would send that exact sequence to $0 \to (0) \to (0) \to \mathbb{Z}/(2) \to 0$, so we lose surjectivity and the Hom functor does not preserve cokernels.

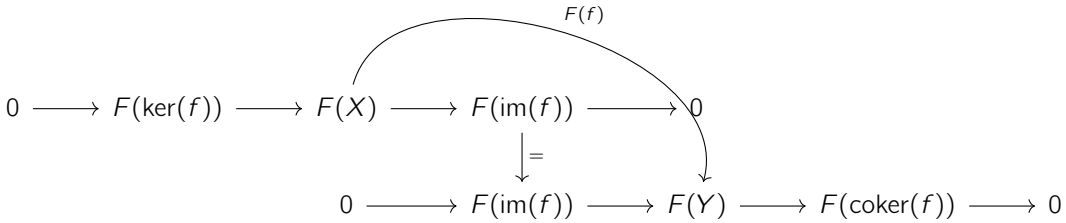
Definition 257

A covariant functor $F : \mathcal{C} \to \mathcal{D}$ is **left exact** (resp. **right exact**) if and only if it preserves kernels (resp. **cokernels**), which is the same as preserving left (resp. right) exact sequences. F is **exact** if it is both left and right exact, meaning that it preserves kernels and cokernels, or equivalently preserving short exact sequences (which is equivalent to preserving all exact sequences).

This last string of logic includes the fact that preserving kernels and cokernels also means we preserve images, and clearly preserving all exact sequences means we preserve short exact sequences. But showing that preserving short exact sequences implies preserving kernels and cokernels takes a bit more work: for any $f : X \to Y$, we can write down two exact sequences as shown below:



Applying our exact functor, we get an exactly analogous diagram where the rows will still be short exact:



We wish to show that $F(\ker(f))$ is the kernel of $F(f)$. Since having $0 \to A \to B$ is the same as having a monomorphism $A \to B$, $F(\operatorname{im}(f)) \to F(Y)$ is a monomorphism. Consider any $Z \to F(X) \xrightarrow{F(f)} F(Y)$ which composes

to the zero map – this factors to a map $Z \to F(X) \xrightarrow{F(f)} F(\text{im}(f)) \to F(Y)$, so $Z \to FX \xrightarrow{F(f)} F(\text{Im}(f))$ is the zero map (because of the monomorphism). That means that $F(\ker f)$ is indeed the kernel of $F(X) \to F(\text{im}(f))$ because the composite map in the top row is zero. The cokernel is checked similarly.

For contravariant functors we have to choose our convention, since a left exact sequence becomes a right exact sequence and vice versa:

Definition 258

Let $F : \mathcal{C} \to \mathcal{D}$ be a contravariant functor. Then F is **left exact**, **right exact**, or **exact**, respectively, if the covariant functor $F^{\text{op}} : \mathcal{C}^{\text{op}} \to \mathcal{D}$ is correspondingly left exact, right exact, or exact.

In other words, F is left exact if and only if for all $X \to Y \to Z \to 0$ right exact, we end up with a left exact sequence in the **target** space $0 \to FZ \to FY \to FX$. (And put another way, left exactness for contravariant functors means that we take cokernels to kernels.)

Example 259

In R-**mod**, the localization functor sending $M \mapsto D^{-1}M$ is exact, the tensor product functor $M \mapsto M \otimes_R N$ (for fixed N) is right exact (but not exact), and the Hom functor $M \mapsto \text{Hom}_R(N, M)$ (for fixed N) is left exact (but not exact). Finally, the contravariant functor $M \mapsto \text{Hom}_R(M, N)$ is left exact (but not exact) as well. The failure of exactness here leads us to the Tor and Ext functors.

Example 260

In **Sh**(X) (the category of sheaves), the global section functor $\mathcal{F} \mapsto \mathcal{F}(X)$ does preserve kernels but not cokernels (by the reasoning in Fact 253), so it is left exact but not generally right exact. Failure of exactness here leads to the usual cohomology theories.

Definition 261

In the category Γ-**Mod** (as defined last lecture), we can send M to M^Γ, the set of fixed points $\{m \in M : \gamma m \; \forall \gamma \in \Gamma\}$; this turns out to be left exact but not generally right exact. And the failure of exactness here turns out to be group cohomology.

Lemma 262

If $0 \to M \to N \to F_R(X) \to 0$ is a short exact sequence of R-modules, and $F : R$-**mod** $\to \mathcal{D}$ is an additive functor, then $0 \to F(M) \to F(N) \to F(F_R(X)) \to 0$ is also exact.

Proof. By the universal property of the free module, if we have a map $0 \to M \xrightarrow{f} N \xrightarrow{g} F_R(X) \to 0$, there exists some $s : F_R(X) \to N$ with $g \circ s = \text{id}$. (Indeed, for each basis element $e_i \in F_R(x)$, we can choose $s(e_i)$ to be any preimage of e_x in g and extend by linearity.) We call s a **section** – notice that $s \circ g$ is not necessarily the identity. Then N is isomorphic to $M \oplus F_R(x)$ by sending $(m, p) \mapsto f(m) + s(p)$, because for any n we can map it to $(f^{-1}(n - s(g(n))), g(n))$ since $g(n - s(g(n))) = 0$, and we can check that these maps are mutually inverses.

So now we also have an analogous sequence $0 \to M \to M \oplus F_R(X) \to F_R(X) \to 0$, such that the following diagram commutes:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M & \xrightarrow{f} & N & \xrightarrow{g} & F_R(X) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle =} & & \downarrow{\scriptstyle \cong} & & \downarrow{\scriptstyle =} & & \\ 0 & \longrightarrow & M & \xrightarrow{\iota_1} & M \oplus F_R(X) & \xrightarrow{\pi_2} & F_R(X) & \longrightarrow & 0 \end{array}$$

Applying the additive functor F and noting that F preserves direct sums, the top and bottom rows are still isomorphic and the bottom row is still short exact, so we still have a short exact sequence $0 \to F(M) \to F(N) \to F(F_R(X)) \to 0$, as desired. □

There's no notion of being a "free module" in a general abelian category, but what we really needed was the section s, so we make definitions that allow for that:

Definition 263

An object $P \in \text{ob}(\mathcal{C})$ is **projective** if for any epimorphism $X \to Y$, any map $P \to Y$ factors through X.

(This is similar to the universal property of a free module – free R-modules are projective in R-**mod**.) There is a dual notion as well:

Definition 264

An object $I \in \text{ob}(\mathcal{C})$ is **injective** if for any monomorphism $X \to Y$, any map $X \to I$ factors through Y.

Lemma 265

If $0 \to X \to Y \to Z \to 0$ is short exact in $\mathcal{C}$, and $F : \mathcal{C} \to \mathcal{D}$ is additive, then $0 \to F(X) \to F(Y) \to F(Z) \to 0$ is short exact in $\mathcal{D}$ if either X is injective or Z is projective.

(This argument is basically the same as the one in the module case – if we have X injective then we construct a map $Y \to X$, and if we have Z projective then we construct a map $Z \to Y$.) The idea is that replacing objects with a collection of projective or injective objects is often useful in homological algebra:

Definition 266

A category $\mathcal{C}$ has **enough projectives** if for any $X \in \text{ob}(\mathcal{C})$, there is an epimorphism $P \to X$ (so any object is a quotient of a projective object). Similarly, $\mathcal{C}$ has **enough injectives** if for any $X \in \text{ob}(\mathcal{C})$ there is some map $X \to I$ with I injective.

We can then iterate this: once we get a map $P^0 \to X$, the kernel of that will be the image of some $P^{-1} \to P^0$, and then the kernel of that will be the image of $P^{-2} \to P^{-1}$, and so on, giving us an exact sequence $\cdots \to P^{-2} \to P^{-1} \to P^0 \to X \to 0$, which we call a **projective resolution** of X. Similarly, if we have enough injectives, repeatedly looking at cokernels gives us an **injective resolution** $X \to I^0 \to I^1 \to I^2 \to \cdots$. (For example, the category of sheaves has enough injectives but not enough projectives.)

24 November 18, 2022

We'll start with a clarification – last time, we said that an additive category is a category in which there are finite products and coproducts, including the empty ones (so the initial and final objects). We then said that the map from

the initial to final object should be an isomorphism, and so should the map $X \amalg Y \to X \times Y$. This gives us an addition $\mathrm{Hom}_{\mathcal{C}}(X, Y)$, but **we require the additional assumption that** for all $x \in \mathcal{C}$, there is a special element $-\mathrm{Id}_X$, such that $-\mathrm{Id}_X + \mathrm{Id}_X = 0$.

Last time, we defined projectives and injectives – projectives are objects P such that whenever we have a surjection $X \to Y$, any map $P \to Y$ can be lifted to a map $P \to X$, and injectives are objects such that whenever we have an injection $X \to Y$, any map $X \to I$ can be extended to a map $Y \to I$. The concepts of "enough projectives" and "enough injectives" are then that we always have some P such that $P \to X$ is an epimorphism, and that we always have some I such that $X \to I$ is a monomorphism, respectively. (In particular, the category R-**mod** has enough projectives and enough injectives, and so does Γ-**mod**, but **Sh**(X) usually has enough injectives but not enough projectives.) Those two conditions give us long exact sequences $\cdots \to P^{-1} \to P^0 \to X \to 0$ and $X \to I^0 \to I^1 \to \cdots$, respectively (by repeatedly applying the condition to the kernel or cokernel of our maps, respectively). Today, we'll discuss injectives and projectives in R-**mod**, and we'll see the use of those projective and injective resolutions next week.

Lemma 267

Let P be an R-module. The following are equivalent:

1. P is projective,
2. P is a direct summand of a free module, meaning that we can write $F_R(\Omega) \cong P \oplus Q$ for some Q.

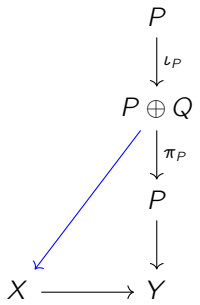
If P is **finitely presented**, meaning that there is a right exact sequence $R^{\oplus b} \to R^{\oplus a} \to P \to 0$, then these are also equivalent to these additional conditions:

3. $P_{\mathfrak{p}}$ is free over $R_{\mathfrak{p}}$ for all prime ideals $\mathfrak{p}$,
4. $P_{\mathfrak{m}}$ is free over $R_{\mathfrak{m}}$ for all maximal ideals $\mathfrak{m}$.

Being **finitely presented** basically means that we want the kernel of the map $R^{\oplus a} \to P$ to also be finitely generated, and one way this is satisfied is if P is finitely generated and R is noetherian.

Proof. To show that (1) implies (2), we know that there is a natural surjection $\pi : F_R(P) \to P$, sending each generator to the corresponding element of P. The property of being projective then means that the identity map $P \to P$ lifts to a map $s : P \to F_R(P)$ such that $\pi \circ s = \mathrm{Id}_P$, so we have the isomorphism $F_R(P) \cong \ker(\pi) \oplus P$ (where in the backward direction we send (a, b) to $a + s(b)$, and in the forward direction we send c to $(c - s\pi(c), \pi(c))$), so P is a direct summand.

For (2) implies (1), we know that $P \oplus Q \cong F_R(\Omega)$. We want to show that if $X \to Y$ is a surjection, then $P \to Y$ lifts to a map $P \to X$. But $P \oplus Q \cong F_R(\Omega)$ is a free module (always projective), so given the composite map $P \oplus Q \xrightarrow{\pi_P} P \to Y$, we get a map $P \oplus Q \to X$ (in blue). Then P maps into $P \oplus Q$ via inclusion in the first factor, and then $\iota_P \circ \pi_P$ is the identity. Thus the composite map $P \xrightarrow{\iota_P} P \oplus Q \to X$ gives us the desired map, showing that P is projective.



Showing equivalence to (3) and (4) is a bit more tricky. Assuming (1) and (2), since P is finitely generated, there is a surjection $F_R(\Omega) \to P$, where $|\Omega|$ is finite. We then get a splitting $s : P \to F_R(\Omega)$ which allows us to write $F_R(\Omega) \cong P \oplus Q$, meaning P is a direct summand of a **finite** free module. Then for any $\mathfrak{p}$, we know that localizing at that prime ideal gives

$$F_{R_\mathfrak{p}}(\Omega) \cong P_\mathfrak{p} \oplus Q_\mathfrak{p}.$$

We then find that (tensoring by $R/\mathfrak{p}$)

$$F_{R_\mathfrak{p}/\mathfrak{p}}(\Omega) \cong P_\mathfrak{p}/\mathfrak{p}P_\mathfrak{p} \oplus Q_\mathfrak{p}/\mathfrak{p}Q_\mathfrak{p},$$

and $(R_\mathfrak{p}/\mathfrak{p})$ is a field so these are finite-dimensional vector spaces. Let the basis elements on the right-hand side be $\overline{e}_1, \cdots \overline{e}_a$ in the first summand and $\overline{e}_{a+1}, \cdots, \overline{e}_b$ in the second. If we choose $e_1, \cdots, e_a \in P_\mathfrak{p}$ that lift $\overline{e}_1, \cdots, \overline{e}_a$ and $e_{a+1}, \cdots, e_b$ in $Q_\mathfrak{p}$ lifting $\overline{e}_{a+1}, \cdots, \overline{e}_b$, by Nakayama's lemma (specifically the corollary which allows us to omit generators in $\mathfrak{p}P_\mathfrak{p}$), we get a surjection $R_\mathfrak{p}^{\oplus a} \to P_\mathfrak{p}$ and a surjection $R_\mathfrak{p}^{\oplus(b-a)} \to Q_\mathfrak{p}$, so we get a surjection $R_\mathfrak{p}^{\oplus b} \to P_\mathfrak{p} \oplus Q_\mathfrak{p} = F_{R_\mathfrak{p}}(\Omega)$. Thus $b = |\Omega|$ and finite-dimensional vector spaces have a fixed basis size, and in particular this map $R_\mathfrak{p}^{\oplus b} \to F_{R_\mathfrak{p}}(\Omega)$ is represented by some $b \times b$ matrix. Reducing that matrix modulo the prime ideal gives us an isomorphism, so $\det(A)$ is nonzero mod $\mathfrak{p}$. But that's the same as saying that $\det(A)$ is not in $\mathfrak{p}$, and $\mathfrak{p}$ is the maximal ideal in $R_\mathfrak{p}$ so $\det(A)$ is a unit. Thus A is invertible, meaning $R_\mathfrak{p}^{\oplus b} \to F_{R_\mathfrak{p}}(\Omega)$ is injective so $R_\mathfrak{p}^{\oplus a} \to P_\mathfrak{p}$ is also injective and thus an isomorphism, as desired.

Clearly (3) implies (4) (since all maximal ideals are prime), and now we prove that (4) implies (1). Given a surjection $M \to N$ and a map $P \to N$, we wish to find a map $P \to M$. We have a map $\mathrm{Hom}_R(P, M) \to \mathrm{Hom}_R(P, N)$, and if we define Q to be its cokernel we have a right exact sequence $\mathrm{Hom}_R(P, M) \to \mathrm{Hom}_R(P, N) \to Q \to 0$. We now localize this at $\mathfrak{m}$ (which preserves exactness) to get $\mathrm{Hom}_R(P, M)_\mathfrak{m} \to \mathrm{Hom}_R(P, N)_\mathfrak{m} \to Q_\mathfrak{m} \to 0$. And we claim that if P is finitely presented, then this can be rewritten as $\mathrm{Hom}_{R_\mathfrak{m}}(P_\mathfrak{m}, M_\mathfrak{m}) \to \mathrm{Hom}_{R_\mathfrak{m}}(P_\mathfrak{m}, N_\mathfrak{m}) \to Q_\mathfrak{m} \to 0$. But since $P_\mathfrak{m}$ is free by assumption, it is certainly projective, so each $Q_\mathfrak{m}$ is zero. And we proved that if $Q_\mathfrak{m} = 0$ for all maximal ideals that means $Q = 0$. So it remains to prove the following assertion:

Lemma 268

Let M and N be R-modules and M be finitely presented. If $D \subset R$ is multiplicative, then the map

$$D^{-1}\mathrm{Hom}_R(M, N) \to \mathrm{Hom}_{D^{-1}(R)}(D^{-1}M, D^{-1}N)$$

is an isomorphism.

We won't go through the full proof, but the idea is to first deal with the case where M is finite and free. Then homomorphisms from $M \to N$ are just a-vectors, so the Hom set is isomorphic to $N^{\oplus a}$. Thus we have a map $D^{-1}N^{\oplus a} \to (D^{-1}N)^{\oplus a}$, which is an isomorphism because localization commutes with direct sums.

Next, if we have a general finitely presented M, meaning we have $R^{\oplus a} \to R^{\oplus b} \to M \to 0$, then applying $\mathrm{Hom}(, N)$ (which is left exact contravariant) yields $0 \to \mathrm{Hom}(M, N) \to \mathrm{Hom}(R^{\oplus b}, N) \to \mathrm{Hom}(R^{\oplus a}, N)$; localizing then gives $0 \to D^{-1}\mathrm{Hom}(M, N) \to D^{-1}\mathrm{Hom}(R^{\oplus b}, N) \to D^{-1}\mathrm{Hom}(R^{\oplus a}, N)$. But we can also localize first and then take Homs, which will yield $0 \to \mathrm{Hom}(D^{-1}M, D^{-1}N) \to \mathrm{Hom}(D^{-1}R^{\oplus b}, D^{-1}N) \to \mathrm{Hom}(D^{-1}R^{\oplus a}, D^{-1}N)$. Putting those into a commutative diagram with two rows, since we know that the maps between the last two columns are isomorphisms, so is the one between the two modules we care about. □

We next talk about injectives, starting with a special case:

Lemma 269

A $\mathbb{Z}$-module I is injective if and only if it is **divisible**, meaning that for all $m \in I$ and $a \in \mathbb{Z}_{\neq 0}$, there is some $m' \in I$ such that $am' = m$.

For example, $\mathbb{Q}$ is a divisible $\mathbb{Z}$-module, as is $\mathbb{Q}/\mathbb{Z}$, so we have a few examples of injective abelian groups.

Proof. If I is injective, then for any $m \in I$ and $a \in \mathbb{Z}_{\neq 0}$ we look at the map $\mathbb{Z} \to I$ sending 1 to m. Then $\mathbb{Z}$ injects into itself by multiplication by a, so by injectivity there is a map (from the latter copy of $\mathbb{Z}$ to M) sending $1 \to m'$, and for the maps to be compatible we must have $m = am'$.

The other direction is a Zorn's lemma argument: if I is divisible and we have an inclusion of abelian groups $X \to Y$ and a map $X \to I$, we consider extending X to Y "bit by bit." Specifically, consider the set $\mathcal{X}$ of pairs (Z, β) where $X \subseteq Z \subset Y$ and $\beta : Z \to I$ is an extension of α (meaning $\beta|_X = \alpha$). This is nonempty because it contains (X, α), and we can place a partial ordering on it where $(Z, \beta) \geq (Z', \beta')$ if $Z \supseteq Z'$ and $\beta|_{Z'} = \beta'$. But if we have a chain of elements $\{(Z, \beta)\}$, it has an upper bound where we take the union W of the submodules Z (still a submodule) and define a map $\gamma : W \to I$ by having $\gamma(w)$ agree with $\beta(w)$ if $w \in Z$ for some (Z, β) (this is consistent because everything is nested).

So by Zorn's lemma, there is some (Z, β) maximal in this set $\mathcal{X}$. If $Z = Y$ we're happy; otherwise, choose some $y \in Y \setminus Z$. We will try to extend β to the module generated by Z and y. Consider the set of multiples of y in Z, which we write as $J = \{n \in \mathbb{Z} : ny \in \mathbb{Z}\}$. This is an ideal of $\mathbb{Z}$ and thus generated by some a. In I, we know that $\beta(ay) = am$ for some $m \in I$ (because I is divisible), and now it makes sense to define $\gamma : \langle Z, y \rangle \to I$ to send $z + by \mapsto \beta(z) + bm$. It remains to check that this is well-defined: indeed, if $z + by = z' + b'y$, then $(z - z') = (b' - b)y$ is an element of Z, which means $b' - b = ca$ for some c. Then the images of the two sides agree because

$$(\beta(z) + bm) - (\beta(z') - b'm) = \beta(z - z') - cam = \beta(z - z' - c(ay)) = \beta((z + by) - (z' + b'y)) = \beta(0) = 0.$$

This contracts maximality and thus $Z = Y$, meaning we've extended our map in the desired way. □

Corollary 270

If I is an injective $\mathbb{Z}$-module and M is a submodule of I, then I/M is also injective (because divisibility is preserved under quotients).

Corollary 271

$\mathbb{Z}$-**mod** has enough injectives, because given a $\mathbb{Z}$-module M we have a short exact sequence $0 \to K \to F_{\mathbb{Z}}(M) \to M \to 0$. In other words, $F_{\mathbb{Z}}(M)/K$ is isomorphic to M. But $F_{\mathbb{Q}}(M)$ is a free $\mathbb{Q}$-module that contains $F_{\mathbb{Z}}(M)$, so M sits inside $F_{\mathbb{Q}}(M)/K$ – since $F_{\mathbb{Q}}(M)$ is divisible (it's a $\mathbb{Q}$-vector space), so is its quotient $F_{\mathbb{Q}}(M)/K$.

This result can be bootstrapped to a general R:

Lemma 272

We have the following:

1. If I is an injective $\mathbb{Z}$-module, then $\mathrm{Hom}_{\mathbb{Z}}(R, I)$ as an R-module (defined by saying that for any $a \in R$ and $f \in \mathrm{Hom}_{\mathbb{Z}}(R, I)$, af is the map $(af)(b) = f(ab)$) is an injective R-module.
2. Thus R-**mod** has enough injectives.

The idea is that $I \to \mathrm{Hom}_{\mathbb{Z}}(R, I)$ is right adjoint to the forgetful functor, and right adjoints preserve injectivity. And given any R-module M, M embeds as an abelian group in some I, so we can embed $M \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, I)$ to get an injective R-module.

25 November 28, 2022

Last time, we showed that there are enough projectives and injectives in R-**mod**, giving a concrete example. We'll now talk about homological algebra properly on more general abelian categories with that in mind:

Lemma 273

Suppose $\mathcal{C}$ is an abelian category with enough injectives. Given any morphism $f : X \to Y$ in $\mathcal{C}$, suppose $0 \to X \to I^0 \to I^1 \to \cdots$ and $0 \to Y \to J^0 \to J^1 \to \cdots$ are injective resolutions (that is, exact sequences with all I^ns and J^ns injective). We also denote these injective resolutions $0 \to X \to I^\bullet$ and $0 \to Y \to J^\bullet$. Then there is a **chain map** (map of complexes) $f : I^\bullet \to J^\bullet$, unique up to **homotopy** (defined below).

If we have two complexes (where remember that we don't need exactness, just that the composite of any two adjacent maps is zero), a chain map between them is a collection of maps $C^i \to D^i$ (vertical maps shown below) such that all squares commute:

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C^{i-1} & \longrightarrow & C^i & \longrightarrow & C^{i+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots & \longrightarrow & D^{i-1} & \longrightarrow & D^i & \longrightarrow & D^{i+1} & \longrightarrow & \cdots
\end{array}$$

Then two maps of complexes $f^\bullet, g^\bullet : C^\bullet \to D^\bullet$ are **homotopic** (denoted $f^\bullet \simeq g^\bullet$) if there are maps $k^i : C^{i+1} \to D^i$ such that $f^i - g^i = \delta_D^{i-1} \circ k^{i-1} + k^i \circ \delta_C^i$ as in the diagram below:

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C^{i-1} & \longrightarrow & C^i & \xrightarrow{\delta_C^i} & C^{i+1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle f^{i-1}-g^{i-1}} & \swarrow{\scriptstyle k^{i-1}} & \downarrow{\scriptstyle f^i-g^i} & \swarrow{\scriptstyle k^i} & \downarrow{\scriptstyle f^{i+1}-g^{i+1}} & & \\
\cdots & \longrightarrow & D^{i-1} & \xrightarrow[\delta_D^{i-1}]{} & D^i & \longrightarrow & D^{i+1} & \longrightarrow & \cdots
\end{array}$$

So the point is that once we replace X and Y with injective objects, we get a map between the extensions, and the extension is unique up to some complicated definition. (The case for enough projectives holds similarly, but we'll leave that to our imagination.)

Proof. We wish to construct $f^0 : I^0 \to J^0$, but J^0 is injective and we have an injection from $X \to I^0$, so there exists a map $f^0 : I^0 \to J^0$ extending the composite map $X \to X \to Y \to J^0$. To extend at degree $i+1$, notice that we can map from I^i to I^{i+1} by mapping $I^i \to \mathrm{coker}\, \delta_I^{i-1} \to I^{i+1}$, and the latter of these is injective (this is just encoding exactness). Now the map $I^{i-1} \to I^i \to J^i \to J^{i+1}$ is the same as the map $I^{i-1} \to J^{i-1} \to J^i \to J^{i+1}$ (by the chain map property), so it is zero. Thus it factors through $\mathrm{coker}\, \delta_I^{i-1}$, and now by injectivity of J^{i+1} the map $f^{i+1} : I^{i+1} \to J^{i+1}$ exists.

For uniqueness, we start with the following diagram:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & X & \longrightarrow & I^0 & \longrightarrow & I^0/X & \longrightarrow & I^1 \\
 & & & & \downarrow{\scriptstyle f^0-g^0} & \swarrow & & \nearrow & \\
0 & \longrightarrow & Y & \longrightarrow & J^0 & & & &
\end{array}$$

Now the map $X \to I^0 \xrightarrow{f^0-g^0} J^0$ is zero, so it factors through I^0/X. Then because $I^0/X \to I_1$ is an injection, we can define the map $k^0 : I^1 \to J^0$ by injectivity of J^0. Then $f^0 - g^0 = k^0 \circ \delta_I^0$ (there's no k^{-1} map so this is what we want at this first stage).

For the general construction of k^i, consider the following diagram:

$$\begin{array}{ccccc} I^{i-1} & \longrightarrow & I^i & \longrightarrow & I^{i+1} \\ {\scriptstyle f^{i-1}-g^{i-1}}\downarrow & \swarrow_{k^{i-1}} & {\scriptstyle f^i-g^i}\downarrow & & \\ J^{i-1} & \longrightarrow & J^i & & \end{array}$$

Then we again split the map $I^i \to I^{i+1}$ into two parts as $I^i \to \operatorname{coker}(\delta_I^{i-1}) \hookrightarrow I^{i+1}$, and we claim the map $f^i - g^i - \delta_J^{i-1} \circ k^{i-1}$ factors through that cokernel. To do so we must check that it's zero on the image of I^{i-1}, but

$$(f^i - g^i - \delta_J^{i-1} \circ k^{i-1}) \circ \delta_I^{l-1} = \delta_J^{i-1} \circ (f^{i-1} - g^{i-1} - k^{i-1} \circ \delta_I^{i-1}) = \delta_J^{i-1} \circ (\delta_J^{i-2} \circ k^{i-2})$$

where we've inductively used the chain property from $i-1$. Then this is indeed zero, and then by injectivity of J^i and the fact that $\operatorname{coker}(\delta_I^{i-1}) \to I^{i+1}$ is an injection we get an extension $k^i : I^{i+1} \to J^i$ which satisfies the desired homotopy property. □

The point in this proof is that once we've come up with the right notion of homotopy, there's only one way the argument can really proceed.

Corollary 274

Let $0 \to X \to I^\bullet$ and $0 \to X \to J^\bullet$ be two injective resolutions of X. Then there is a (unique up to homotopy) $f^\bullet : I^\bullet \to J^\bullet$ extending the identity map Id_X and also $g^\bullet : J^\bullet \to I^\bullet$ extending Id_X, Then $g^\bullet \circ f^\bullet$ is a map of complexes from $I^\bullet \to I^\bullet$ extending the identity, so $g^\bullet \circ f^\bullet$ is homotopic to the identity map on the complexes. The same holds for $f^\bullet \circ g^\bullet$.

In general we say that two complexes $I^\bullet$ and $J^\bullet$ are **homotopic** if there are maps $f^\bullet : I^\bullet \to J^\bullet$ and $g^\bullet : J^\bullet \to I^\bullet$ such that the compositions in both directions are homotopic to their respective identities. So now that we've replaced an object with a complex of injectives up to homotopy, we might want to ask what properties are preserved under homotopy. The answer is **cohomology**:

Definition 275

Let $C^\bullet$ be a complex, not necessarily exact. The ith **(co)homology** $H^i(C^\bullet)$ is defined as follows: we have maps $C^{i-1} \xrightarrow{\delta_C^{i-1}} C^i \xrightarrow{\delta_C^i} C^{i+1}$, and we know that $\operatorname{im} \delta_C^{i-1}$ is contained in $\ker \delta_C^i$, so we may set $H^i(C^\bullet) = \ker \delta_C^i / \operatorname{im} \delta_C^{i-1}$.

For example, if $0 \to X \to I^\bullet$ is an injective resolution, then $H^i(I^\bullet)$ is zero for all i except at $i = 0$ by exactness. And at $i = 0$, the map $I^0 \to I^1$ has a kernel, specifically X.

This cohomology has nice functorial properties: if $f : C^\bullet \to D^\bullet$ is a map of complexes, we get a map $H^i : H^i(C^\bullet) \to H^i(D^\bullet)$ because f^i maps the image of δ_C^{i-1} to the image of δ_D^{i-1} and also the kernel of δ_C^i to the kernel of δ_D^i. So it maps the quotient appropriately as well.

Lemma 276

If $f^\bullet$ and $g^\bullet$ are two homotopic maps $C^\bullet \to D^\bullet$, then $H^i(f^\bullet) = H^i(g^\bullet)$ as maps $H^i(C^\bullet) \to H^i(D^\bullet)$.

Proof. We'll do the proof for just R-modules. For any $x \in \ker \delta_C^i$, we know that we map to $f^i(x)$ and $g^i(x)$ under the two maps of complexes, but

$$f^i(x) - g^i(x) = (\delta_D^{i-1} \circ k^{i-1} + k^i \circ \delta_C^i)(x) = \delta_D^{i-1}(k^{i-1}(x)),$$

which is the image of δ_D^{i-1}. Since cohomology mods out that image, $f^i(x) = g^i(x)$ will be the same under the cohomology maps. □

Definition 277

Let $\mathcal{C}$ and $\mathcal{D}$ be abelian categories. Suppose $\mathcal{C}$ has enough injectives, and suppose $F : \mathcal{C} \to \mathcal{D}$ is a left exact, additive functor. Then if $X \in \text{ob}(\mathcal{C})$, then it has an injective resolution $0 \to X \to I^\bullet$. Then $FI^\bullet$ is a complex (though not necessarily exact anymore), and we can look at its cohomology. We can then define the ***i*th right derived functor** of F via

$$R^iF(X) = H^i(FI^\bullet)$$

This appears to depend on the injective resolution, but it turns out it does not: if $0 \to X \to I^\bullet$ and $0 \to X \to J^\bullet$, then we get an extension of the identity map $f^\bullet : I^\bullet \to J^\bullet$, unique up to homotopy, such that $f^\bullet \circ g^\bullet \simeq \text{Id}_J$ and $g^\bullet \circ f^\bullet \simeq \text{Id}_I$. Then applying $Fg^\bullet : FJ^\bullet \to FI^\bullet$ and $Ff^\bullet : FI^\bullet \to FJ^\bullet$, we know that $Ff^\bullet \circ Fg^\bullet \simeq \text{Id}_{FJ^\bullet}$ because F preserves homotopy. (This is a general property of additive functors: $f^\bullet \simeq_{k^\bullet} g^\bullet$, then $f^i - g^i = \delta^{i-1} \circ k^{i-1} + k^i \circ \delta^i$, so $F(f^i) - F(g^i) = F(\delta^{i-1}) \circ F(k^{i-1}) + F(k^i) \circ F(\delta^i)$, meaning that $F(f^\bullet) \simeq F(g^\bullet)$.) So $Ff^\bullet \circ Fg^\bullet$ is homotopic to the identity on $FJ^\bullet$, and similarly $Fg^\bullet \circ Ff^\bullet$ is homotopic to the identity on $FI^\bullet$, so $H^i(Ff^\bullet) \circ H^i(Fg^\bullet)$ is the identity map and so is $H^i(Fg^\bullet) \circ H^i(Ff^\bullet)$. And if $f'^\bullet$ also extended the identity map, then $f^\bullet$ is homotopic to $f'^\bullet$, so $Ff^\bullet$ and $Ff'^\bullet$ are homotopic and thus the **same** map on cohomology. So it's not just unique up to isomorphism – it's unique up to **unique** isomorphism.

So now if we have a map $f : X \to Y$, we can choose resolutions $I^\bullet$ and $J^\bullet$ and extend f on the resolutions to get well-defined maps $H^i(Ff^\bullet) : H^i(FI^\bullet) \to H^i(FJ^\bullet)$, and that is how we define $R^iF(f)$. So we know how objects and morphisms are sent under R^iF, and we can check that R^iF is an additive functor (identity and composition are preserved). And R^0F can be described explicitly: if F is left exact, then $0 \to X \to I^0 \to I^1$ being left exact means $0 \to FX \to FI^0 \to FI^1$ is also left exact, so in fact $R^0F = F$ (since we compute this zeroth cohomology by looking at the kernel of $FI^0 \to FI^1$). And if I is injective, $R^iF(I) = (0)$ for all higher $i > 0$. Indeed, an injective resolution for I is just $0 \to I \xrightarrow{id} I \to 0 \to 0 \to \cdots$, and the cohomology of $FI \to 0 \to 0 \to \cdots$ is trivial except in $i = 0$.

Example 278

Let X be a topological space, and we have a **global section** functor $\Gamma : \mathbf{Sh}(X) \to \mathbf{Ab}$ sending $\mathcal{F}$ to the abelian group $\mathcal{F}(X)$ (basically setting U the entire space). This is left exact, and $\mathbf{Sh}(X)$ has enough injective, so we get functors $R^i\Gamma : \mathbf{Sh}(X) \to \mathbf{Ab}$ for all $i \geq 0$.

Theorem 279

Suppose X is a second countable topological space (meaning it has a countable base), and suppose every point has an open neighborhood homeomorphic to an open set in $\mathbb{R}^n$ (for example any manifold). Then the derived functors of the global section functor can be applied to the constant sheaf $\mathbb{Z}$. We then have

$$R^i\Gamma(\mathbb{Z}_X) \cong H^i_{\text{sing}}(X, \mathbb{Z})$$

where $H^i_{\text{sing}}(X)$ is the usual singular cohomology of X with coefficients in $\mathbb{Z}$ from algebraic topology.

26 November 30, 2022

Last time, we considered the following situation: if $\mathcal{C}, \mathcal{D}$ are abelian categories with $\mathcal{C}$ having enough injectives, and we have an additive functor $F : \mathcal{C} \to \mathcal{D}$ which is covariant and left exact, then we can define the right derived (covariant) functors $R^iF : \mathcal{C} \to \mathcal{D}$ in the following way: any object X has an injective resolution $0 \to X \to I^\bullet$, and throwing away X and then applying F gives us a sequence $FI^\bullet$ (no longer necessarily exact, but still a complex). Then we set $R^iF(x) = H^i(FI^\bullet)$; we showed last time that this is indeed functorial and well-defined. (And in particular, $R^0F = F$.)

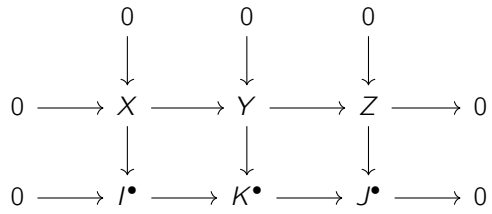
We can similarly see that if F is **contravariant** but still left exact, and $\mathcal{C}$ has enough **projectives**, we can again define (now contravariant) right derived functors $R^iF : \mathcal{C} \to \mathcal{D}$. Here, we take a projective resolution $P^\bullet \to X \to 0$, throw away X, and apply F to get a complex $FP^\bullet$. (In other words, given $P^{-2} \to P^{-1} \to P^0 \to X \to 0$, we look at the complex $FP^0 \to FP^{-1} \to FP^{-2} \to \cdots$. Just to keep notation consistent, we define $(FP)^i = FP^{-i}$, and then we can define $R^iF(X) = H^i((FP)^\bullet)$ – all of the analogous properties will still hold.

The two other similar situations also work out the way we might expect – if F is covariant and **right exact**, and $\mathcal{C}$ has enough **projectives**, we now get (covariant) **left derived functors** $L^iF : \mathcal{C} \to \mathcal{D}$, in which we take a projective resolution $P^\bullet \to X \to 0$, throw away X and apply F to get $FP^\bullet$, and define $L^iF(X) = H^{-i}(FP^\bullet)$. (So for example, L^1F is the kernel of the map $FP^{-1} \to FP^0$, modulo the image of the map $FP^{-2} \to FP^{-1}$.) And finally, if F is **contravariant** and **right exact**, and $\mathcal{C}$ has enough injectives, we get (contravariant) left derived functors $L^iF : \mathcal{C} \to \mathcal{D}$ by taking the injective resolution $0 \to X \to I^\bullet$, throw away X and apply F to get $FI^\bullet$. From $I^0 \to I^1 \to I^2 \to \cdots$ we thus get a complex $\cdots \to FI^2 \to FI^1 \to F^0$, and we can again renumber and define $((FI)^\bullet)^i = F^{-i}$ so that $L^iF(X) = H^{-i}((FI)^\bullet)$.

To say a bit more about these left and right derived functors, they were introduced because the original functors F are not fully exact. Motivated by that, if we have $0 \to X \to Y \to Z \to 0$, and X, Y, Z have injective resolutions, we may want to relate the short exact sequence on X, Y, Z to short exact sequences simultaneously on all terms of the injective resolutions.

Proposition 280

Let $\mathcal{C}$ have enough injectives, and suppose $0 \to X \to Y \to Z \to 0$ is short exact in $\mathcal{C}$. Then there exist injective resolutions $0 \to X \to I^\bullet, 0 \to Y \to K^\bullet, 0 \to Z \to J^\bullet$, such that the diagram below commutes and the rows are exact:



(Like last time, there is an exactly analogous situation for projectives.)

Proof. We've shown previously that $0 \to I \to Y \to Z \to 0$ is short exact with I injective, then we in fact have $Y \cong I \oplus \mathbb{Z}$. So the only option for proving this result is if we can show $K^i \cong I^i \oplus J^i$. Thus we can choose injective resolutions $0 \to X \to I^\bullet$ and $0 \to Z \to J^\bullet$ and set $K^i = I^i \oplus J^i$. (So notice that in fact this result is **stronger** than stated – we can choose any $I^\bullet$ and $J^\bullet$ and there is still a suitable K.)

We'll do the proof in the case of R-modules. Our goal is to show that $0 \to Y \to K^\bullet$ is an injective resolution, and we must define maps $Y \to I^0 \oplus J^0$ and $I^i \oplus J^i \to I^{i+1} \oplus J^{i+1}$. Since we need the square with Y, Z, K^0, J^0 to commute, we must send y to $(h^{-1}y, \delta_J^{-1}y)$, such that (now looking at the square X, Y, I^0, K^0) h^{-1} restricted to X is δ_I^{-1}. (Here note that the -1s are **indices**, not inverses.) And such a map h^{-1} exists because I^0 is injective, so a map $X \to I^0$ extends to Y through our map $X \to Y$. So that gives us $0 \to I^0 \to I^0 \oplus J^0 \to J^0 \to 0$, so the diagram we want commutes with $i = 0$. And now we can check that $Y \to I^0 \oplus J^0$ is injective because mapping to zero requires us to be zero by going around the right square.

For a general stage i, we know the map from $I^i \oplus J^i$ should send (x, y) to (something, $\delta_J^i y$) by commutativity of the right square, and in fact we must have $(x, y) \mapsto (\delta_I^i x + h^i y, \delta_J^i y)$ where h^i is a map $J^i \to I^{i+1}$. It remains to check that this process gives an injective resolution $0 \to Y \to I^0 \oplus J^0 \to I^1 \oplus J^1 \to \cdots$ if we choose h^i correctly, so we must check exactness. First of all, we must have $\delta \circ \delta = 0$, so exactness at $I^{i+1} \oplus J^{i+1}$ yields

$$(x, y) \mapsto (\delta_I^i x + h^i y, \delta_I^i y) \mapsto (\delta_I^{i+1} h^i y + h^{i+1} \delta_J^i y, 0)$$

and thus we need $\boxed{\delta_I^{i+1} h^i + h^{i+1} \delta_J^i = 0}$. (And we also need to think about the case $i = -1$ separately, but in that case it turns out we do require $\delta_I^0 h^{-1} + h^0 \delta_J^{-1} = 0$.) But it turns out this condition **also** guarantees exactness – if $(x, y) \in I^i \oplus J^i$ maps to zero in $I^{i+1} \oplus J^{i+1}$, then $\delta_J^i y = 0$ and $\delta_I^i x + h^i y = 0$, but by exactness of J we know $y = \delta_J^{i-1} y'$ for some y', and thus the second equation becomes $0 = \delta_I^i x + h^i \delta_J^{i-1} y' = \delta_I^i (x - h^{i-1} y')$ (last step by our condition). Then by exactness of I we see that $x + h^{i-1} y' = \delta_I^{i-1} x'$ for some x', so in fact (x', y') maps to $(\delta_I^{i-1} x' + h^{i-1} y', \delta_J^{i-1} y') = (x, y)$. So exactness is automatic and we just need to construct h^i satisfying that boxed condition.

And we can do this inductively – we've already constructed h^{-1}, and the construction of h^0 is left as an exercise (it's similar). For the general case, we want to construct a map $h^{i+1} : J^{i+1} \to I^{i+2}$ for any $i \geq 0$ which makes the map $J^i \to I^{i+2} : -\delta_I^{i+1} \circ h^i$ factor through $J^i \to J^{i+1}$. But we can first map $J^i \to \operatorname{coker}(\delta_I^{i-1})$, and we can check that $J^i \to I^{i+2}$ factors through that cokernel. Indeed, the composite map $J^{i-1} \to J^i \to I^{i+2}$ is $-\delta_I^{i+1} h^i \delta_J^{i-1}$, and by inductive hypothesis we can use the condition for i to show that this is in fact $\delta_I^{i+1} \delta_I^i h^{i-1} = 0$. So we get a map $\operatorname{coker}(\delta_I^{i-1}) \to I^{i+2}$, and then by injectivity of I^{i+2} this gives us a map $J^{i+1} \to I^{i+2}$ because $\operatorname{coker}(\delta_I^{i-1}) \to J^{i+1}$ is injective. □

So now we know that we can choose compatible $I^\bullet, K^\bullet, J^\bullet$ injective resolutions – applying F to them, we get an exact sequence $0 \to FI^\bullet \to FK^\bullet \to FJ^\bullet \to 0$. What we've shown is that the columns (each of the complexes) will no longer be exact, but the rows are exact (so we lose exactness $FX \to FY \to FZ$, but we still have exactness $FI^\bullet \to FK^\bullet \to FJ^\bullet$).

Lemma 281

Suppose $C^\bullet, D^\bullet, E^\bullet$ are complexes that fit into a commutative diagram $0 \to C^\bullet \xrightarrow{f^\bullet} D^\bullet \xrightarrow{g^\bullet} E^\bullet \to 0$ with exact rows. By functoriality, we get an exact sequence $H^i(C^\bullet) \to H^i(D^\bullet) \to H^i(E^\bullet)$, and it turns out we can extend this to a sequence $H^i(C^\bullet) \to H^i(D^\bullet) \to H^i(E^\bullet) \to H^{i+1}(C^\bullet) \to H^{i+1}(D^\bullet) \to H^{i+1}(E^\bullet) \to \cdots$ which is exact everywhere. (The blue map is called the **boundary map**.)

Corollary 282

If $F : \mathcal{C} \to \mathcal{D}$ is left exact, $\mathcal{C}$ has enough injectives, and we have $0 \to X \to Y \to Z \to 0$ exact in $\mathcal{C}$, then we get an exact sequence

$$0 \to R^0FX \to R^0FY \to R^0FZ \to R^1FX \to R^1FY \to R^1FZ \to R^2FX \to \cdots,$$

and because this sequence starts as $0 \to FX \to FY \to FZ$ we have a way of measuring the failure of exactness under F.

Beginning of proof of Lemma 281. Consider the diagram below:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & C^{i-1} & \longrightarrow & D^{i-1} & \longrightarrow & E^{i-1} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & C^{i} & \longrightarrow & D^{i} & \longrightarrow & E^{i+1} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & C^{i+1} & \longrightarrow & D^{i+1} & \longrightarrow & E^{i+1} & \longrightarrow & 0
\end{array}$$

The kernel of the map $\delta_C^i : C^i \to C^{i+1}$ maps to the kernel of the map δ_D^i, which maps to the kernel of δ_E^i (since the image of something in C^i going to zero in C^{i+1} goes to zero in D^{i+1}, and so on). Furthermore, the composite map $\ker \delta_C^i \to \ker \delta_D^i \to \ker \delta_E^i$ is zero because the composite map $C^i \to D^i \to E^i$ is zero. We also similarly have a map $\operatorname{im}(\delta_C^{i-1}) \to \operatorname{im}(\delta_D^{i-1}) \to \operatorname{im}(\delta_D^i)$. So to show exactness at $H^i(D^\bullet)$, we start with some $d \in \ker \delta_D^I$ and consider what happens to $d + \operatorname{im}(\delta_D^{i-1})$. We have $g^i(d) = \delta_E^{i-1}(e) = \delta_E^{i-1}(g^{i-1}(d')) = g^i(\delta_D^{i-1}(d'))$ – thus, $g^i(d - \delta_D^{i-1}d') = 0$, so $d - \delta_D^{i-1}d' = f^i(c)$ for some c. But by injectivity, to show $\delta_C^i(c) = 0$ it suffices to compute $f^{i+1}(\delta_C^i(c)) = \delta_D^i(f^i(c)) = \delta_D^i(d) - \delta_D^i\delta_D^{i-1}d' = 0$. So $\delta_D^i(i) = 0$, and now for any $c + \operatorname{im}(\delta_C^{i-1})$, we map onto $f^i(c) + \operatorname{im}(\delta_D^{I-1})$. So in summary starting with something in $H^i(D)$ (specifically $d + \operatorname{im}\delta_D^{i-1})$) which maps under g to zero, we found some $c + \operatorname{im}(\delta_C^{i-1})$ which maps to it. That shows exactness at $H^i(D^\bullet)$, and next time we'll show exactness at $H^i(C^\bullet)$ and $H^i(E^\bullet)$ as well.

□

27 December 2, 2022

Last lecture, we mentioned that given an exact sequence of complexes $0 \to C^\bullet \to D^\bullet \to E^\bullet \to 0$, there are maps in cohomology $H^i(C^\bullet) \to H^i(D^\bullet) \to H^i(E^\bullet)$ for each i, and we can in fact add boundary maps $H^i(E^\bullet) \to H^{i+1}(C^\bullet)$ to make an exact sequence $\cdots \to H^i(C^\bullet) \to H^i(D^\bullet) \to H^i(E^\bullet) \to H^{i+1}(C^\bullet) \to H^{i+1}(D^\bullet) \to H^{i+1}(E^\bullet) \to \cdots$.

Proof of Lemma 281, continued. We look at the diagram from last time again:

$$\begin{array}{ccccccccc}
 & & & & D^{i-1} & \xrightarrow{g^{i-1}} & E^{i-1} & \longrightarrow & 0 \\
 & & & & \downarrow{\scriptstyle \partial_D^{i-1}} & & \downarrow{\scriptstyle \partial_E^{i-1}} & & \\
0 & \longrightarrow & C^i & \xrightarrow{f^i} & D^i & \xrightarrow{g^i} & E^i & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \partial_C^i} & & \downarrow{\scriptstyle \partial_D^i} & & \downarrow{\scriptstyle \partial_E^i} & & \\
0 & \longrightarrow & C^{i+1} & \xrightarrow{f^{i+1}} & D^{i+1} & \xrightarrow{g^{i+1}} & E^{i+1} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \partial_C^{i+1}} & & \downarrow{\scriptstyle \partial_D^{i+1}} & & & & \\
0 & \longrightarrow & C^{i+2} & \longrightarrow & D^{i+2} & & & &
\end{array}$$

To construct a map $E^i \to C^{i+1}$, we start with some element $e \in \ker \partial_E^i$ and want to consider where $e + \mathrm{im}(\partial_E^{i-1})$ goes. By surjectivity, there is some $d \in D^k$ so that $g^i d = e$. Then we know that $g^{i+1}(\partial_D^i d) = \partial_E^i e = 0$ (by going around the bottom right square in both ways, and by using the definition of e), so by exactness there is some $c \in C^{i+1}$ such that $f^{i+1}c = \partial_D^i d$. But we know that $\partial_C^{i+1}c$ will map to an element in D^{i+2}, but the map $C^{i+2} \to D^{i+2}$ is injective and $\partial_D^{i+1}f^{i+1}c = \partial_D^{i+1}\partial_D^i d = 0$, which means c is in the kernel of ∂_C^{i+1}. We thus **define** the boundary map by sending $e + \mathrm{im}(\partial_E^{i-1})$ to $c + \mathrm{im}(\partial_C^i)$.

We must check that this is well-defined, since we made arbitrary choices e and d in this construction. Indeed, if $e' + \mathrm{im}(\partial_E^{i-1}) = e + \mathrm{im}(\partial_E^{i-1})$, then $e' - e = \partial_E^{i-1}e''$ for some $e'' \in E^{i-1}$. Lifting e'' to $d'' \in D^{i-1}$ (since the map $D^{i-1} \to E^{i-1}$ is surjective), we have $g^{i-1}d'' = e''$, so

$$e' = e\partial_E^{i-1}e'' = \partial_E^{i-1}g^{i-1}d'' = g^i\partial_D^{i-1}d''$$

(by looking at the top right square), and thus $d' - d - \partial_D^{i-1}d'' \in \ker(g^i)$, so by exactness in the ith row we can write it as $f^i c''$ for some $c'' \in C^i$. Applying ∂_D^i to both sides, $\partial_D^i(d') - \partial_D^i(d) = \partial_D^i\partial_D^{i-1}d'' + \partial_D^i f^i c''$ – the first term on the right-hand side is zero, and the second term becomes $f^{i+1}\partial_C^i c''$. And now remembering that we defined $f^{i+1}(c') = \partial_D^i(d')$, we see that

$$f^{i+1}(c') - f^{i+1}(c) = f^{i+1}\partial_C^i(c'') \implies c' - c = \partial_C^i c''$$

by injectivity of f^{i+1}. But this means that $c' + \mathrm{im}(\partial_C^i) = c + \mathrm{im}(\partial_C^i)$, so regardless of our choice of d and e we end up with the same element in cohomology, meaning our map $H^i(E^\bullet) \to H^{i+1}(C^\bullet)$ is well-defined.

We now need to check that we have a complex: first, we check that the composite map $H^i(D^\bullet) \to H^i(E^\bullet) \to H^{i+1}(C^\bullet)$ is zero. A typical element of $H^i(D^\bullet)$ is of the form $d + \mathrm{im}(\partial_D^{i-1})$ where $\partial_D^i d = 0$; this is then sent to $g^i(d) + \mathrm{im}(\partial_E^{i-1})$. But then following the prescription of our boundary map, we first find a preimage of $g^i(d)$, and an obvious choice is d itself; by definition we are then sent to $\partial_D^i d + \mathrm{im}(\partial_C^i)$ (viewed as an element of $C^{i+1} \subset D^{i+1}$). But the composite map must then be zero because $\partial_D^i d = 0$ by definition. Similarly, the composite $e + \mathrm{im}(\partial_E^{i-1}) \mapsto c + \mathrm{im}(\partial_C^i) \mapsto f^{i+1}c + \mathrm{im}(\partial_C^i)$ is zero, because $f^{i+1}c$ is in $\mathrm{im}(\partial_D^i)$ by definition.

Next, we must check exactness itself, first at $H^i(E^\bullet)$ – suppose $e + \mathrm{im}(\partial_E^{i-1})$ maps to 0 in $H^{i+1}(C^\bullet)$. That means that c would be an an element of $\mathrm{im}(\partial_C^i)$, so $c = \partial_C^i c'$ for some $c' \in C^i$. We wish to show that this means e is in the image of $g^i : D^i \to E^i$ – indeed, $\partial_D^i d = f^{i+1}c = f^{i+1}\partial_C^i c' = \partial_D^i f^i c'$, so $d - f^i c'$ is in the kernel of ∂_D^i. Then applying g^i to $(d - f^i c') + \mathrm{im}(\partial_D^{i-1}) \in H^i(D^\bullet)$, we see that

$$(d - f^i c') + \mathrm{im}(\partial_D^{i-1}) \mapsto g^i d - g^i f^i c' + \mathrm{im}(\partial_E^{i-1}),$$

but $g^i f^i$ is zero by exactness of row i, and $g^i d = e$ by definition, meaning that e is indeed mapped to by some element in $H^i(D^\bullet)$. Finally, to check exactness at $H^i(C^\bullet)$, start with some $c + \mathrm{im}(\partial_C^i) \in H^{i+1}(C^\bullet)$ such that $f^{i+1}c \in \mathrm{im}(\partial_D^i)$; we wish to prove this is mapped to by the boundary map. If we say that $f^{i+1}c = \partial_D^i d$, then we can consider

$g^i d + \mathrm{im}(\partial_E^{i-1}) \in H^i(E^\bullet)$, and under our boundary map we are mapped to $c + \mathrm{im}(\partial_C^i) \in H^i(C^\bullet)$, as desired. □

We've done the proof here just for R-modules, but the argument works for general abelian categories – it just looks messier because we can't talk about elements in the same way.

Lemma 283

Suppose there is a commutative diagram of complexes with exact rows as shown below:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & C^\bullet & \longrightarrow & D^\bullet & \longrightarrow & E^\bullet & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f^\bullet} & & \downarrow{\scriptstyle g^\bullet} & & \downarrow{\scriptstyle h^\bullet} & & \\ 0 & \longrightarrow & C'^\bullet & \longrightarrow & D'^\bullet & \longrightarrow & E'^\bullet & \longrightarrow & 0 \end{array}$$

Then the diagram below also commutes:

$$\begin{array}{ccccccccccc} \cdots & \longrightarrow & H^i(C^\bullet) & \longrightarrow & H^i(D^\bullet) & \longrightarrow & H^i(E^\bullet) & \longrightarrow & H^{i+1}(C^\bullet) & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle H^i(f^i)} & & \downarrow{\scriptstyle H^i(g^i)} & & \downarrow{\scriptstyle H^i(h^i)} & & \downarrow{\scriptstyle H^{i+1}(f^{i+1})} & & \\ \cdots & \longrightarrow & H^i(C'^\bullet) & \longrightarrow & H^i(D'^\bullet) & \longrightarrow & H^i(E'^\bullet) & \longrightarrow & H^{i+1}(C'^\bullet) & \longrightarrow & \cdots \end{array}$$

We basically just need to check that the new square formed by the boundary map commutes – this is a similar argument to what we just did, and it's left as an exercise to us.

Recall that if we have an additive left-exact functor $F : \mathcal{C} \to \mathcal{D}$ with $\mathcal{C}$ having enough injectives, then $0 \to X \to Y \to Z \to 0$ gives us a long exact sequence $0 \to R^0FX \to R^0FY \to R^0FZ \to R^1FX \to \cdots$ coming from the exact sequence of injective resolutions $0 \to FI^\bullet \to FK^\bullet \to FJ^\bullet \to 0$ (here using injectivity).

Lemma 284

Suppose $\mathcal{C}, \mathcal{D}$ are abelian categories, $F : \mathcal{C} \to \mathcal{D}$ is left exact and additive, and $\mathcal{C}$ has enough injectives. Then given a commutative diagram with exact rows between $0 \to X \to Y \to Z \to 0$ and $0 \to X' \to Y' \to Z' \to 0$, we get the following commutative diagram:

$$\begin{array}{ccccccccccccc} 0 & \longrightarrow & R^0FX & \longrightarrow & R^0FY & \longrightarrow & R^0FZ & \longrightarrow & R^1FX & \longrightarrow & R^1FY & \longrightarrow & R^1FZ & \longrightarrow & R^2FX & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle R^0Ff} & & \downarrow{\scriptstyle R^0Fg} & & \downarrow{\scriptstyle R^0Fh} & & \downarrow{\scriptstyle R^1Ff} & & \downarrow{\scriptstyle R^1Fg} & & \downarrow{\scriptstyle R^1Fg} & & \downarrow{\scriptstyle R^2Ff} & & \\ 0 & \longrightarrow & R^0FX' & \longrightarrow & R^0FY' & \longrightarrow & R^0FZ' & \longrightarrow & R^1FX' & \longrightarrow & R^1FY' & \longrightarrow & R^1FZ' & \longrightarrow & R^2FX' & \longrightarrow & \cdots \end{array}$$

Proof sketch. Using the proof last time, we can find injective resolutions $I^\bullet, I^\bullet \oplus J^\bullet, J^\bullet$ of X, Y, Z so that $0 \to I^\bullet \to I^\bullet \oplus J^\bullet \to J^\bullet \to 0$ fit into the commutative diagram, using the maps $k^i : J^i \to I^{i+1}$, and then having the map $I^i \oplus J^i \to I^{i+1} \to J^{i+1}$ send $(x, y) \to (\partial_I x + k^i y, \partial_J^i y)$. Then we want to apply our result about long exact sequences to $0 \to FI^\bullet \to F(I^\bullet \oplus J^\bullet) \to FJ^\bullet \to 0$. But we can compare the results we get from $0 \to X \to Y \to Z \to 0$ and $0 \to X' \to Y' \to Z' \to 0$. Given a map $f : X \to X'$, we saw that the injective resolutions give us a map $f^\bullet : I^\bullet \to I'^\bullet$. Similarly, for a map $h : Z \to Z'$, we get a map $h^\bullet : J^\bullet \to J'^\bullet$. So now we want to try to take our map $g : Y \to Y'$ and get a map $f^\bullet \oplus h^\bullet$ from $I^\bullet \oplus J^\bullet \to I'^\bullet \oplus J'^\bullet$. We want to send (x, y) to $(\partial_{I'} x + \ell^i y, \partial_{J'}^\bullet y)$, where we choose $\ell^i : J^i \to J'^{(i+1)}$ to satisfy the properties that **(1)** $0 \to Y' \to I'^\bullet \oplus J'^\bullet$ is a complex, and **(2)** $f^\bullet \oplus g^\bullet : I^\bullet \oplus J^\bullet \to I'^\bullet \oplus J'^\bullet$ is a map of complexes. (So the key point is that we do everything on $0 \to X \to Y \to Z \to 0$ first, and then we must construct the map ℓ^i similarly to how we construct k^i but now with an extra condition coming from compatibility – it's important that this is the **last** thing we do.) □

These right derived functors we've defined are part of a more general notion:

Definition 285

A **δ-functor** $\mathcal{C} \to \mathcal{D}$, denoted $\{S^\bullet\}$, is a sequence of additive functors $S^0, S^1, S^2, \cdots$ all from $\mathcal{C}$ to $\mathcal{D}$, such that for any $0 \to X \to Y \to Z \to 0$ exact in $\mathcal{C}$, we get a long exact sequence $0 \to S^0X \to S^0Y \to S^0Z \to S^1X \to S^1Y \to S^1Z \to \cdots$ (so these boundary maps $S^0Z \to S^1X$ are part of the definition of the δ-functor). Furthermore, the "extra square" formed from functoriality must commute, meaning that for any commutative diagram with exact rows $0 \to X \to Y \to Z \to 0$ and $0 \to X' \to Y' \to Z' \to 0$, we must have commutativity within the square formed by $S^iZ, S^{i+1}X, S^iZ'$, and $S^{i+1}X'$.

This definition is a bit hard to work with, but there's an extra property we may have that makes working with these objects easier:

Definition 286

A δ-functor $\{S^\bullet\}$ is a **universal δ-functor** if for any other δ-functor $\{T^\bullet\}$ and any natural transformation $\phi^0 : S^0 \to T^0$, there is a unique natural transformation $\phi^i : S^i \to T^i$ for all i such that the square formed by $S^iZ, S^{i+1}X, T^iZ$, and $T^{i+1}X$ commutes for any exact sequence $0 \to X \to Y \to Z \to 0$.

So the point is that we often understand the map at degree 0 quite well, and that can help us get maps between the higher degrees without the construction. And it turns out that R^iF is a universal δ-functor – this is a useful technique called **dimension shifting**, and we'll go through the proof of that next time.

28 December 5, 2022

Last time, we considered the following situation: if $\mathcal{C}$ has enough injectives and $F : \mathcal{C} \to \mathcal{D}$ is a left exact additive functor, we defined the right derived functors $R^iF : \mathcal{C} \to \mathcal{D}$ by taking an injective resolution $0 \to X \to I^\bullet$, applying F to it, and taking its cohomology. Canonically this doesn't depend on the choice of injective resolution, and it is indeed a functor. Then if $0 \to X \to Y \to Z \to 0$ is exact, we found that we get a long exact sequence $R^iFX \to R^iFY \to R^iFZ \to R^{i+1}FX \to \cdots$ (and in fact this is functorial). A generalization of this idea is a δ-functor, which is a sequence of functors $\{S^i\}$ such that such a long exact sequence (including the boundary maps) can be constructed. Specifically, it's useful to understand **universal δ-functors**, which are δ-functors $\{S^i\}$ where given a natural transformation $\phi^0 : S^0 \to T^0$, we can produce $\phi^i : S^i \to T^i$ for all i in a way that is compatible with exact sequences.

Lemma 287

The sequence $\{R^iF\}$ is a universal δ-functor.

Proof. We've already seen that this sequence is a δ-functor, so we just need to check universality. Let S^i be another δ-functor, and suppose $\phi^0 : R^0F = F \to S^0$ is a natural transformation. We will inductively construct $\phi^i : R^iF \to S^i$, prove it is a natural transformation, and show that it is compatible with boundary maps. Basically, given any $X \in \text{ob}(\mathcal{C})$, we must construct $\phi^i_X : R^iF(X) \to S^i(X)$. The way we do this is by embedding X in a sequence $0 \to X \to I \to Q \to 0$ for some injective I (where Q is the resulting quotient). We then get a sequence $R^{i-1}FI \to R^{i-1}FQ \to R^iFX$ as part of our long exact sequence. But we showed that the higher derived functors R^iFI vanish for injectives I, so in fact the map $R^{i-1}FQ \to R^iFX$ is surjective. We also have $S^{i-1}I \to S^{i-1}Q \to S^iX$ as part of the long exact sequence for S,

so putting these together we can fill in this part of the diagram, with left square commuting because ϕ_I^{i-1} is a natural transformation (but dashed map not yet constructed)

$$\begin{array}{ccccccc} R^{i-1}FI & \longrightarrow & R^{i-1}FQ & \longrightarrow & R^iFX & \longrightarrow & 0 \\ \downarrow{\scriptstyle \phi_I^{i-1}} & & \downarrow{\scriptstyle \phi_Q^{i-1}} & & \vdots{\scriptstyle \phi_X^i} & & \\ S^{i-1}FI & \longrightarrow & S^{i-1}FQ & \longrightarrow & S^iFX & \longrightarrow & 0 \end{array}$$

Since R^iFX is the cokernel of the map $R^{i-1}FI \to R^{i-1}FQ$, we can factor the map $R^{i-1}FQ \to S^iFX$ through R^iFX as long as it is zero on $\mathrm{im}(R^{i-1}FI)$. And indeed, the map $R^{i-1}FI \to R^{I_1}FQ \to S^{i-1}Q \to S^{i-1}X$ is zero because it's the same as traveling along $R^{i-1}FI \to S^{i-1}FI \to S^{i-1}FQ \to S^iFX$, and the composition of those last two maps is zero by exactness. So we do factor through R^iFX in a unique way, and now we need to check that this is independent of I and that this is indeed a natural transformation. We can do those together by considering an arbitrary map $f : X \to Y$ and considering exact sequences $0 \to X \to I \to Q \to 0$ and $0 \to Y \to J \to Q' \to 0$ (the point is that checking independence of I can be done by taking $X = Y$ but choosing different injective resolutions). Then we have a diagram as below (except the blue dashed arrows):

$$\begin{array}{ccccccccc} 0 & \longrightarrow & X & \longrightarrow & I & \longrightarrow & Q & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \vdots{\scriptstyle g} & & \vdots{\scriptstyle h} & & \\ 0 & \longrightarrow & Y & \longrightarrow & J & \longrightarrow & Q' & \longrightarrow & 0 \end{array}$$

J is injective, so the monomorphism $X \to J$ factors through I, meaning we can draw $g : I \to J$ to make the left square commute. Then again Q is the cokernel of $X \to I$, so to produce a map $h : Q \to Q'$ we just need to check that the map $X \to I \to J \to Q'$ is zero, which it is by exactness of the bottom row. So now we need to check the natural transformation condition (and remember that setting f to be the identity will show that our definition of ϕ_X^i does not depend on whether we use I or J as the injective resolution for X):

$$\begin{array}{ccc} R^iFX & \xrightarrow{R^iFf} & R^iFY \\ \downarrow{\scriptstyle \phi_{X,I}^i} & & \downarrow{\scriptstyle \phi_{Y,J}^i} \\ S^iFX & \xrightarrow{S^iFf} & S^iFY \end{array}$$

For that, we first look at this commutative diagram, which commutes because ϕ_I^{i-1} is natural:

$$\begin{array}{ccc} R^{i-1}FQ & \xrightarrow{R^{i-1}Fh} & R^{i-1}FQ' \\ \downarrow{\scriptstyle \phi_Q^{i-1}} & & \downarrow{\scriptstyle \phi_{Q'}^{i-1}} \\ S^{i-1}FQ & \xrightarrow{S^iFh} & S^{i-1}FQ' \end{array}$$

Then there is a surjective map from the top left corner here ($R^{i-1}FQ$) to the top left corner R^iFX in the previous diagram, as well as a map from top right corner $R^{I-1}FQ'$ to top right corner R^iFY. Because $R^{i-1}F$ is a δ-functor, the square formed by those four elements commute. Similarly, we get a commutative square from the four elements on the bottom. So we know those faces of our "cube" commute, and the front and back square commute because of the definition of $\phi_{X,I}^i$ and $\phi_{Y,J}^i$. So basically everything commutes except the original diagram we drew, and now we can check that that original face also commutes because we can start from some element R^iFX and pull it back to $R^{i-1}FQ$ (by surjectivity); then we can basically follow the two paths to S^iY by using commutativity of all of the other faces. So our maps are well-defined and natural transformations.

Next, we check that this is compatible with boundary maps. For that, look at a short exact sequence $0 \to X \to Y \to Z \to 0$, yielding a boundary map $R^{i-1}FZ \to R^iFX$ and also one for $S^{i-1}FZ \to S^iFX$:

$$\begin{array}{ccc} R^{i-1}FZ & \longrightarrow & R^iFX \\ \downarrow{\scriptstyle \phi_Z^{i-1}} & & \downarrow{\scriptstyle \phi_X^i} \\ S^{i-1}FZ & \longrightarrow & S^iFX \end{array}$$

From our previous construction, we know this square already commutes when Y is injective. So we'll map X into some injective, and we know the corresponding square for $0 \to X \to I \to Q \to 0$ commutes. But $X \to Y$ is an injection so it extends to a map $Y \to I$, and Q is a cokernel so we get a map $Z \to Q$:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & X & \longrightarrow & Y & \longrightarrow & Z & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & X & \longrightarrow & I & \longrightarrow & Q & \longrightarrow & 0 \end{array}$$

Thus we get a commutative square by definition of ϕ^i:

$$\begin{array}{ccc} R^{i-1}FQ & \longrightarrow & R^iFX \\ \downarrow{\scriptstyle \phi_Q^{i-1}} & & \downarrow{\scriptstyle \phi_X^i} \\ S^{i-1}FQ & \longrightarrow & S^iFX \end{array}$$

We compare those two squares. The square formed by the four "right elements" commutes because the two S^iXs and R^iFXs are equal. The map $Z \to Q$ induces maps $R^{i-1}FZ \to R^{i-1}FQ$ and $S^{i-1}Z \to S^{i-1}Q$, so again we form a cube. Then because ϕ^{i-1} is natural, the square formed by the four "left elements" also commutes. Also the four top elements, as well as the four bottom elements, commute because R^i and S are δ-functors and thus commutes with boundary maps. So again all faces of the cube commute except the original one, and this is true because the map $S^iX \to S^iX$ is an injection (in fact an equality), so we can chase the arrows $R^{i-1}FZ \to R^iFX \to S^iX \to S^iX$ and $R^{i-1}FZ \to S^{i-1}Z \to S^iX \to S^iX$ around both ways and find that we indeed get the same map. □

The whole principle of this proof is that we reduced something in degree i to something in degree $(i-1)$ by considering injectives instead.

We'll next make another general definition:

Definition 288

An object $X \in \text{ob}(\mathcal{C})$ is **acyclic** for an additive functor F if $R^iFX = 0$ for all $i > 0$.

For example, we've seen that all injective objects are acyclic, but sometimes there are acyclic objects that are not injective and it's useful to work with the more general class (at least for the study of the functor F alone).

Lemma 289

If X embeds into some acyclic object A via $0 \to X \to A \to Q \to 0$, then we get maps $FA \to FQ \to R^1FX \to 0$ and $0 \to R^iFQ \to R^{i+1}FX \to 0$ for all $i > 0$.

In other words, we can understand the higher right derived functors applied to X in terms of how they are applied to the corresponding quotient one degree down. And we can repeat this process as well:

Lemma 290

Suppose we have a sequence of objects Q_i and a sequence of acyclic objects A_i, such that $Q_0 = X$ and $0 \to Q_i \to A_i \to Q^{i+1} \to 0$ is short exact for all i. Then applying the previous argument, we have

$$R^m FX \cong R^{m-1} FQ_1 \cong R^{m-2} FQ_2 \cong \cdots \cong R^1 FQ_{m-1} = \text{coker}(FA_{m-1} \to FQ_m).$$

So if we can embed in this way, we can understand the higher derived functors in terms of the cokernel of F applied to a single map. And we can do even better: if $0 \to Q_m \to A_m \to Q_{m+1} \to 0$ is our sequence and we apply F to it, we get $0 \to FQ_m \to FA_m \to FQ_{m+1}$ because F is left exact. Then Q_{m+1} embeds into A_{m+1}, so looking $Q_m \to A_m \to A_{m+1}$ we still have left exactness $0 \to FQ_m \to FA_m \to FA_{m+1}$, so $FQ_m = \ker(FA_m \to FA_{m+1})$. Plugging this into the previous fact, we actually find that

$$R^m FX \cong R^m(FA^\bullet),$$

since this is basically saying that we have an acyclic resolution $0 \to X \to A_0 \to A_1 \to A_2 \to \cdots$ of X. So that brings us back to the idea that we had with injective resolutions.

Our last topic of the class will be the **Ext** and **Tor** functors, which are somehow the simplest applications of homological algebra to describe but not the most interesting. (But it appears regularly on the qualifying exam.)

We'll begin with Ext – suppose $\mathcal{C}$ is a general abelian category. Fixing any $X \in \text{ob}(\mathcal{C})$, we know that $\text{Hom}_{\mathcal{C}}(X, \cdot)$ is a covariant, left exact functor from $\mathcal{C}$ to **Ab**. We can then think about the right derived functors of it:

Definition 291

If $\mathcal{C}$ has enough injectives, we define

$$\overline{\text{Ext}}^i_{\mathcal{C}}(X, \cdot) = R^i \text{Hom}(X, \cdot).$$

Similarly, if Y is any object of $\mathcal{C}$, we have a contravariant functor $\text{Hom}_{\mathcal{C}}(\cdot, Y) : \mathcal{C} \to$ **Ab** which will again be left exact. We can then look at its right derived functors as well:

Definition 292

If $\mathcal{C}$ has enough projectives, we define

$$\text{Ext}^i_{\mathcal{C}}(\cdot, Y) = R^i \text{Hom}_{\mathcal{C}}(\cdot, Y).$$

So more concretely, if $0 \to Y_1 \to Y_2 \to Y_3 \to 0$ is a short exact sequence, then we get a map

$$0 \to \text{Hom}_{\mathcal{C}}(X, Y_1) \to \text{Hom}_{\mathcal{C}}(X, Y_2) \to \text{Hom}_{\mathcal{C}}(X, Y_3) \to \overline{\text{Ext}}^1(X, Y_1) \to \overline{\text{Ext}}^1(X, Y_2) \to \cdots,$$

and similarly if we have a short exact sequence $0 \to X_1 \to X_2 \to X_3 \to 0$, we get a map

$$0 \to \text{Hom}_{\mathcal{C}}(X_3, Y) \to \text{Hom}_{\mathcal{C}}(X_2, Y) \to \text{Hom}_{\mathcal{C}}(X_1, Y) \to \text{Ext}^1(X_3, Y) \to \text{Ext}^1(X_2, Y) \to \cdots.$$

But the point is that these are actually the same if they both exist (so measuring lack of exactness will yield the same result whether we are taking homomorphisms into the objects or out of them).

Theorem 293

If $\mathcal{C}$ has enough injectives and enough projectives, then $\text{Ext}^i_{\mathcal{C}}(X, Y) \cong \overline{\text{Ext}}^i_{\mathcal{C}}(X, Y)$.

To show this, we'll need a better characterization of Ext:

Lemma 294

The following are equivalent:

1. X is projective,
2. $\mathrm{Ext}^i(X, Y) = 0$ for all $i > 0$ and all Y,
3. $\mathrm{Ext}^1(X, Y) = 0$ for all Y.

Proof. To show that (1) implies (2), we calculate $\mathrm{Ext}^i(X, Y)$ by finding a projective resolution of X, but we can just use $\cdots \to 0 \to 0 \to X \to X \to 0$ in this case. And we calculate $\mathrm{Ext}^i(X, Y)$ by looking at the cohomology

$$\mathrm{Hom}(X, Y) \to \mathrm{Hom}(0, Y) \to \mathrm{Hom}(0, Y) \to \cdots,$$

which is $\mathrm{Hom}(X, Y)$ if $i = 0$ and 0 otherwise, so indeed condition (2) is satisfied. (2) clearly implies (3). Finally, suppose (3) holds. Given X we can find some projective P such that we get a sequence $0 \to K \to P \to X \to 0$ – we claim that this sequence splits, so that X is a direct sum of a projective and thus projective itself. For that, we need to construct a map $P \to K$ by looking at homomorphisms into K: there is an exact sequence $0 \to \mathrm{Hom}(X, K) \to \mathrm{Hom}(P, K) \to \mathrm{Hom}(K, K) \to \mathrm{Ext}^1(X, K) \to \cdots$, but by assumption $\mathrm{Ext}^1(X, K)$ is zero. Thus the map $\mathrm{Hom}(P, K) \to \mathrm{Hom}(K, K)$ is surjective, meaning some map $f : P \to K$ lifts to the identity map $1_K \in \mathrm{Hom}(K, K)$. That means that the composition of the maps $K \xrightarrow{\iota} P \xrightarrow{f} K$ is the identity (by definition of the Hom functor), so P is isomorphic to $K \oplus \ker f$ and $\ker f$ is isomorphic to X. □

We'll develop some more results on Ext and prove that these two definitions are indeed equivalent next time!

29 December 7, 2022

Last time, we considered the covariant, left exact functor $\mathrm{Hom}_{\mathcal{C}}(X, \cdot)$, which gives us right derived functors $\overline{\mathrm{Ext}}^i_{\mathcal{C}}(X, \cdot)$ if $\mathcal{C}$ has enough injectives, and the contravariant, left exact functor $\mathrm{Hom}_{\mathcal{C}}(\cdot, Y)$, which gives right derived functors $\mathrm{Ext}^i_{\mathcal{C}}(\cdot, Y)$ if $\mathcal{C}$ has enough projectives. Our goal is to show that these two Ext functors are actually the same when they both exist – last time, we showed that (when $\mathcal{C}$ has enough projectives) an object is projective if and only if $\mathrm{Ext}^i(X, Y) = 0$ for all $i > 0$ and Y (that is, X is acyclic for $\mathrm{Hom}_{\mathcal{C}}(\cdot, Y)$ for all Y), and in fact it's also equivalent to just checking that $\mathrm{Ext}^1(X, Y) = 0$ for all Y. We'll now prove a similar criterion for injectivity:

Lemma 295

Suppose $\mathcal{C}$ has enough projectives (we'll assume this throughout the rest of the lecture). Then the following are equivalent:

1. Y is injective,
2. $\mathrm{Ext}^i(X, Y) = 0$ for all $i > 0$ and all X (in other words, $\mathrm{Ext}^i(\cdot, Y) = 0$ for all $i > 0$),
3. $\mathrm{Ext}^1(X, Y) = 0$ for all X.

Proof. For (1) implies (2), start with a projective resolution $P^\bullet \to X \to 0$. We then get $0 \to \mathrm{Hom}(X, Y) \to \mathrm{Hom}(P^\bullet, Y)$, and then throwing away $\mathrm{Hom}(X, Y)$ gives us cohomology. So we're just trying to prove that we have

an exact sequence $\mathrm{Hom}(P^\bullet, Y)$, but $\mathrm{Hom}(\cdot, Y)$ is exact if and only if Y is injective, so we indeed have trivial Ext^i for all $i > 0$ and all X, as desired. (To explain why $\mathrm{Hom}(\cdot, Y)$ is exact, start with $0 \to A \to B \to C \to 0$. Taking homomorphisms into Y, we know it's left exact so we have $0 \to \mathrm{Hom}(C, Y) \to \mathrm{Hom}(B, Y) \to \mathrm{Hom}(A, Y)$, so the question is whether that last map is surjective. So we want to ask whether a map from A to Y is a restriction of a map B to Y, but that's the definition of being injective.)

(2) implies (3) is again clear. For (3) implies (1), suppose we have $0 \to A \to B$, and there is a map $A \to Y$ and we want to show that we can extend it to a map $B \to Y$. Indeed, we can complete the exact sequence to $0 \to A \to B \to Q \to 0$ – the corresponding exact sequence is

$$0 \to \mathrm{Hom}(Q, Y) \to \mathrm{Hom}(B, Y) \to \mathrm{Hom}(A, Y) \to \mathrm{Ext}^1(Q, Y) \to \cdots,$$

and by assumption $\mathrm{Ext}^1(Q, Y) = 0$ so $\mathrm{Hom}(B, Y) \to \mathrm{Hom}(A, Y)$ is indeed surjective, meaning that any map $A \to Y$ can indeed be lifted. □

(Note that in the previous proof, we similarly used the fact that X is projective if and only if $\mathrm{Hom}(X, \cdot)$ is exact.) Next, recall that $\mathrm{Ext}^i(X, Y)$ was defined as a functor in X, but we want to show that it actually behaves as a functor in Y as well:

Lemma 296
The map $Y \mapsto \mathrm{Ext}^i(X, Y)$ is a functor.

Proof. Given a map $f : Y_1 \to Y_2$, we need to produce a map $\mathrm{Ext}^1(X, Y_1) \to \mathrm{Ext}^1(X, Y_2)$. We'll do so by producing a natural transformation $\mathrm{Ext}^i(\cdot, Y_1) \to \mathrm{Ext}^i(\cdot, Y_2)$ (and then we can substitute any X in). Since these are δ-functors, we have a map $\phi : \mathrm{Hom}(\cdot, Y_1) \to \mathrm{Hom}(\cdot, Y_2)$ sending each morphism g to $f \circ g$. We must check that this is natural by taking any $h : X \to X'$ and checking the diagram:

$$\begin{array}{ccc} \mathrm{Hom}(X', Y_1) & \xrightarrow{\phi_{X'}} & \mathrm{Hom}(X', Y_2) \\ \downarrow{\scriptstyle -\circ h} & & \downarrow{\scriptstyle -\circ h} \\ \mathrm{Hom}(X, Y_1) & \xrightarrow{\phi_{X'}} & \mathrm{Hom}(X, Y_2) \end{array}$$

Indeed, going around the bottom takes a morphism from g to $g \circ h$ to $f \circ g \circ h$, and going around the top takes us from g to $f \circ g$ to $f \circ g \circ h$, so we do have a natural transformation. (The other diagram is checked similarly.) And because $\mathrm{Ext}^i(\cdot, Y_2)$ is a δ-functor (also universal, but not important) and $\mathrm{Ext}^i(\cdot, Y_1)$ is a universal δ-functor, for each i there is a unique natural transformation $\phi^i : \mathrm{Ext}^i(\cdot, Y_1) \to \mathrm{Ext}^1(\cdot, Y_2)$ compatible with the boundary maps. So in fact $\mathrm{Ext}^1(X, Y_1) \to \mathrm{Ext}^1(X, Y_2)$ can just be ϕ^1_X – we can then check that functoriality does actually hold because of naturality. □

Lemma 297
For any $0 \to Y \to Z \to W \to 0$, we get a long exact sequence $\mathrm{Ext}^i(X, Z) \to \mathrm{Ext}^i(X, W) \to \mathrm{Ext}^{i+1}(X, Y) \to \mathrm{Ext}^{i+1}(X, Z) \to \cdots$.

Again, we already know we have a long exact sequence in the first variable, but we're saying that we get this δ-functor-like property in the second variable as well.

Proof. The recipe for $\mathrm{Ext}^i(X,Z)$ was to take a projective resolution $P^\bullet \to X \to 0$, take homomorphisms $\mathrm{Hom}(P^\bullet, Y) \to \mathrm{Hom}(P^\bullet, Z) \to \mathrm{Hom}(P^\bullet, W)$. If this sequence were short exact for each P^i, we would get the desired long exact sequence, so we must just check that short exactness holds. But that's true because P^i is projective so $\mathrm{Hom}(P^i, \cdot)$ is exact. □

Lemma 298
$\mathrm{Ext}^i(X, \cdot)$ is a δ-functor.

Proof sketch. We've already constructed the long exact sequence, and we just need to check that it is functorial. But each long exact sequence $0 \to \mathrm{Hom}(P^\bullet, Y) \to \mathrm{Hom}(P^\bullet, Z) \to \mathrm{Hom}(P^\bullet, W) \to 0$ can be thought of as a "plane" with P^is in one direction and Y, Z, W in the other, and we just need to diagram chase with two such planes next to each other. This is left as an exercise to us. □

Lemma 299
Now suppose $\mathcal{C}$ also has enough injectives. We have an isomorphism $\overline{\mathrm{Ext}}^i(X,Y) \to \mathrm{Ext}^i(X,Y)$ – in fact, there is a natural isomorphism $\overline{\mathrm{Ext}}(X,Y) \to \mathrm{Ext}^i(X, \cdot)$.

(This means that $\overline{\mathrm{Ext}}^i(X,Y)$ is isomorphic to $\mathrm{Ext}^i(X,Y)$ in a way functorial in Y.)

Proof. The usual way of constructing maps between right derived functors is as follows: we start with the identity map $\mathrm{Hom}(X, \cdot) \to \mathrm{Hom}(X, \cdot)$, which is a natural transformation. By construction, $\overline{\mathrm{Ext}}^i(X, \cdot)$ is a universal δ-functor, and we've just checked that $\mathrm{Ext}^i(X, \cdot)$ is a δ-functor as well in Lemma 298, though this time we don't know it is universal yet because it didn't arise from a right derived functor. (It turns out that δ-functor that vanishes on injectives must be a universal δ-functor, so we could check that it vanishes on injectives, which is true by the first lemma of this lecture. But we didn't discuss that fact yet.) Thus there are unique maps $\overline{\mathrm{Ext}}^i(X, \cdot) \to \mathrm{Ext}^i(X, \cdot)$ compatible with boundary maps which extend the identity in degree 0.

We will show this is an isomorphism by induction on i using a dimension-shifting argument: since we have enough injectives, we have a short exact sequence $0 \to Y \to I \to Q \to 0$, which yields a long exact sequence $\overline{\mathrm{Ext}}^{i-1}(X, I) \to \overline{\mathrm{Ext}}^{i-1}(X, Q) \to \overline{\mathrm{Ext}}^{i-1}(X, Y) \to \overline{\mathrm{Ext}}^{i}(X, I)$. But because $i > 0$ and I is injective, $\overline{\mathrm{Ext}}^{i}(X, I) = 0$. Similarly, we have the same long exact sequence for Ext^i, namely $\mathrm{Ext}^{i-1}(X, I) \to \mathrm{Ext}^{i-1}(X, Q) \to \mathrm{Ext}^{i-1}(X, Y) \to \mathrm{Ext}^{i}(X, I) = (0)$. We then get maps between them:

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & \overline{\mathrm{Ext}}^{i-1}(X,I) & \longrightarrow & \overline{\mathrm{Ext}}^{i-1}(X,Q) & \longrightarrow & \overline{\mathrm{Ext}}^{i-1}(X,Y) & \longrightarrow & \overline{\mathrm{Ext}}^{i}(X,I) = (0) \longrightarrow \cdots \\
 & & \downarrow \cong & & \downarrow \cong & & \vdots & & \\
\cdots & \longrightarrow & \mathrm{Ext}^{i-1}(X,I) & \longrightarrow & \mathrm{Ext}^{i-1}(X,Q) & \longrightarrow & \mathrm{Ext}^{i-1}(X,Y) & \longrightarrow & \mathrm{Ext}^{i}(X,I) = (0) \longrightarrow \cdots
\end{array}$$

By induction, we have an isomorphism in the first two columns, and we want to check if the third column is an isomorphism. But that's because $\overline{\mathrm{Ext}}^i(X,Y)$ and $\mathrm{Ext}^i(X,Y)$ are both cokernels of isomorphic maps (from the first to the second column in both the top and bottom rows), so they are isomorphic. And in fact when $i > 1$ the first column vanishes as well so we have isomorphisms between Y in degree i and Q in degree $(i-1)$. □

Example 300

We'll work in the example of $\mathbb{Z}$-modules (abelian groups). Notice that $\operatorname{Ext}^i_{\mathbb{Z}}(\mathbb{Z}, A) = A$ if $i = 0$ and 0 otherwise, because $\mathbb{Z}$ is a projective $\mathbb{Z}$-module. On the other hand, we can compute $\operatorname{Ext}^i_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, A)$ by looking at the short exact sequence $0 \to \mathbb{Z} \xrightarrow{\times n} \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \to 0$, yielding a long exact sequence

$$0 \to \operatorname{Hom}(\mathbb{Z}/n\mathbb{Z}, A) \to \operatorname{Hom}(\mathbb{Z}, A) \to \operatorname{Hom}(\mathbb{Z}, A) \to \operatorname{Ext}^1(\mathbb{Z}/n\mathbb{Z}, A) \to \operatorname{Ext}^1(\mathbb{Z}, A),$$

where the last term is in fact zero.

We know that the blue map is induced by the multiplication-by-n map, so $\operatorname{Ext}^1(\mathbb{Z}/n\mathbb{Z}, A)$ is the cokernel of that, which is A/nA, and $\operatorname{Hom}(\mathbb{Z}/n\mathbb{Z}, A)$ is the n-torsion of A, denoted $A[n]$. And then because $\operatorname{Ext}^i(\mathbb{Z}, A) = 0$ for all $i > 0$, we find that $\operatorname{Ext}^i(\mathbb{Z}/n\mathbb{Z}, A) = 0$ if $i > 1$, A/nA if $i = 1$, and $A[n]$ if $i = 0$. Additionally, by functoriality, a short exact sequence $0 \to A \to B \to C \to 0$ yields $0 \to \operatorname{Ext}^0(\mathbb{Z}/n\mathbb{Z}, A) \to \operatorname{Ext}^0(\mathbb{Z}/n\mathbb{Z}, B) \to \operatorname{Ext}^0(\mathbb{Z}/n\mathbb{Z}, C) \to \cdots$, or more specifically an exact sequence

$$0 \to A[n] \to B[n] \to C[n] \to A/nA \to B/nB \to C/nC \to 0 \to 0 \to \cdots.$$

(This is basically the snake lemma if we have multiplication-by-n maps between two copies of $0 \to A \to B \to C \to 0$.)

Example 301

As a challenge, we should try computing $\operatorname{Ext}^i(\mathbb{Q}, \mathbb{Z})$. There's no homomorphisms $\mathbb{Q} \to \mathbb{Z}$, so it turns out this is zero or $i = 0$ or $i > 1$, but there is a surprisingly complicated answer for $i = 1$.

This is all we'll say about Ext, and we're now ready to talk about Tor. We'll just make the definition for now: recall that $M \otimes \cdot$ is a functor R-**mod** $\to$ R-**mod** sending N to $M \otimes N$. This is covariant and right-exact, and R-**mod** has enough projectives. Thus, we produce functors

$$\operatorname{Tor}^R_i(M, \cdot) = L^i(M \otimes_R \cdot)$$

for each i. This means that for any exact sequence $0 \to N_1 \to N_2 \to N_3 \to 0$, we get a long exact sequence (showing failure of left exactness)

$$\cdots \to \operatorname{Tor}^R_1(M, N_1) \to \operatorname{Tor}^R_1(M, N_2) \to \operatorname{Tor}^R_1(M, N_3) \to M \otimes N_1 \to M \otimes N_2 \to M \otimes N_3 \to 0.$$

Definition 302

An R-module M is **flat** if $M \otimes_R \cdot$ is exact.

We'll see that flat R-modules turn out to be the acyclic ones, so we can use flat modules in place of projectives. But we'll understand this more next time.

30 December 9, 2022

Last time, we started discussing Tor: using the right-exact, covariant tensor $M \otimes_R \cdot$ taking R-**mod** to itself, we get the left derived functors $L_i(M \otimes_R \cdot) = \operatorname{Tor}^R_i(M, \cdot)$, and we get the usual long exact sequence from the short exact

sequence. We mentioned that a module M is **flat** if $M \otimes_R \cdot$ is exact (not just right exact), and this immediately gives us the following equivalent definitions (since the derived functors of an exact functor are trivial):

Lemma 303

The following are equivalent:

1. M is a flat R-module,
2. $\mathrm{Tor}_i(M, \cdot) = 0$ for all $i > 0$ (in other words, $\mathrm{Tor}_i(M, N) = 0$ for all $i > 0$ and all modules N),
3. $\mathrm{Tor}_1(M, \cdot) = 0$.

Proof. For (1) implies (2), take a projective resolution $P^\bullet \to N \to 0$. Then $\mathrm{Tor}_i(M, N) = 0$ is the $-i$th cohomology of the complex $H^{-i}(M \otimes P^i)$, but by exactness (because M is flat) tensoring will still give us an exact sequence and thus zero Tor for all $i > 0$.

(2) implies (3) is clear. For (3) implies (1), suppose we have an exact sequence $0 \to N_1 \to N_2 \to N_3 \to 0$. Tensoring with M yields a long exact sequence $\mathrm{Tor}_1(M, N_3) \to M \otimes N_1 \to M \otimes N_2 \to M \otimes N_3 \to 0$, but we assume $\mathrm{Tor}_1(M, N_3) = 0$ so this is indeed short exact. □

Lemma 304

$M_1 \oplus M_2$ is flat if and only if M_1 and M_2 are flat (because tensor products distribute over direct sums). Also, R is a flat R-module, and so is a general free R-module. In particular, projective modules are direct summands of free modules, so they are also flat.

Recall that Ext had the strange property that it can be considered a functor in either variable, and we'll now establish something similar for Tor.

Lemma 305

Given a short exact sequence $0 \to M_1 \to M_2 \to M_3 \to 0$, there is a long exact sequence "the wrong way around:"

$$\cdots \to \mathrm{Tor}_1(M_2, N) \to \mathrm{Tor}_1(M_3, N) \to M_1 \otimes N \to M_2 \otimes N \to M_3 \otimes N \to 0$$

Proof. Take a projective resolution $P^\bullet \to N \to 0$. But since each P^i is projective, it is also flat, which means that the rows of $0 \to M_1 \otimes P^\bullet \to M_2 \otimes P^\bullet \to M_3 \otimes P^\bullet \to 0$ are exact. The long exact sequence arising from this then gives us the desired long exact sequence. □

Lemma 306

The following are equivalent:

1. N is a flat R-module,
2. $\mathrm{Tor}_i(M, N) = 0$ for all $i > 0$ and all M (that is, N is acyclic for the functor $M \otimes_R \cdot$ for all M),
3. $\mathrm{Tor}_1(M, N) = 0$ for all M.

Proof. For (1) implies (2), we will argue by induction on i. For any such M, we can write it as part of a short exact sequence $0 \to K \to P \to M \to 0$ with P projective (and K the kernel of that map $P \to M$). By the previous lemma,

and the fact (the first thing we proved today) that $\mathrm{Tor}_i(P, N) = 0$ because P is projective and thus flat, we get a long exact sequence

$$\mathrm{Tor}_i(P, N) \to \mathrm{Tor}_i(M, N) \to \mathrm{Tor}_{i-1}(K, N) \to \mathrm{Tor}_{i-1}(P, N).$$

If $i = 1$, the last two modules here are instead $K \otimes N \to P \otimes N$, and since K injects in P and N is flat, $K \otimes N$ will also inject into $P \otimes N$, meaning that exactness must yield $\mathrm{Tor}_1(M, N) = 0$. (So this is really our base case.) Otherwise if $i > 1$, we know that $\mathrm{Tor}_{i-1}(K, N) = 0$ by the inductive hypothesis, and we know $\mathrm{Tor}_i(P, N) = 0$ because P is projective and flat. Thus we have exactness $0 \to \mathrm{Tor}_i(M, N) \to 0$ and thus $\mathrm{Tor}_i(M, N) = 0$ as desired.

(2) implies (3) is trivial. For (3) implies (1), we have an exact sequence $0 \to M_1 \to M_2 \to M_3 \to 0$ and we tensor it with N. We then get (the "wrong way around" long exact sequence) the long exact sequence

$$\mathrm{Tor}_1(M_3, N) \to M_1 \otimes N \to M_2 \otimes N \to M_3 \otimes N \to 0,$$

and by assumption the first module here is zero and we do have exactness. □

Proposition 307

The map $\{M \mapsto \mathrm{Tor}_i(M, N)\}$ is a δ-functor ("in the wrong variable").

Proof. To check that this is a functor, we take a morphism $f : M_1 \to M_2$ and get a natural transformation $\phi : M_1 \otimes \cdot \to M_2 \otimes \cdot$, where $\phi_N = f \otimes \mathrm{Id}_N$ sends $M_1 \otimes N \to M_2 \otimes N$ in the usual way. (Naturality can be checked from the definition.) Because the ordinary Tor definition is a universal δ-functor (because it is a deried functor): we get natural transformations $\phi^i : \mathrm{Tor}_i(M_1, \cdot) \to \mathrm{Tor}_i(M_2, \cdot)$ and thus $\phi^i_N : \mathrm{Tor}_i(M_1, N) \to \mathrm{Tor}_i(M_2, N)$.

Alternatively, we could also check more directly: taking a projective resolution $P^\bullet \to N \to 0$, we get a map $M_1 \otimes P^\bullet \to M_2 \otimes P^\bullet$ which gives rise to cohomology $H^i(M_1 \otimes P^\bullet) \to H^i(M_2 \otimes P^\bullet)$.

Finally, by the lemmas we've just proved, we do get long exact sequences, and we need to check that they are functorial – that last part is left as an exercise to us. □

Lemma 308

Tor is symmetric in a natural way – that is, there is an isomorphism $\mathrm{Tor}_i(M, N) \to \mathrm{Tor}_i(N, M)$.

Proof. Fix M. Then $\mathrm{Tor}_i(M, \cdot)$ is a **universal** δ-functor, and $\mathrm{Tor}_i(\cdot, M)$ is a (not-necessarily universal for now) δ-functor. Furthermore, in degree 0 we have a map $M \otimes \cdot \to \cdot \otimes M$ is the usual map sending $m \otimes n \mapsto n \otimes m$, which we can check is a natural transformation.

Thus universality gives us natural transformations $\phi_i : \mathrm{Tor}_i(M, \cdot) \to \mathrm{Tor}_i(\cdot, m)$ compatible with boundary maps, and we want to prove that we have an isomorphism. But like last time, we can use a dimension shifting argument for this. We use induction on the statement "$\phi_{i,N}$ is an isomorphism for all N." As mentioned, the base case $i = 0$ is true because $M \otimes N \to N \otimes M$ is an isomorphism, and now for $i > 0$ we can put N into an exact sequence $0 \to K \to P \to N \to 0$ with P projective. We then get two long exact sequences in the rows

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & \mathrm{Tor}_i(M, P) & \longrightarrow & \mathrm{Tor}_i(M, N) & \longrightarrow & \mathrm{Tor}_{i-1}(M, K) & \longrightarrow & \mathrm{Tor}_{i-1}(M, P) \longrightarrow \cdots \\
 & & & & \downarrow & & \downarrow \cong & & \downarrow \cong \\
\cdots & \longrightarrow & \mathrm{Tor}_i(P, M) & \longrightarrow & \mathrm{Tor}_i(N, M) & \longrightarrow & \mathrm{Tor}_{i-1}(K, M) & \longrightarrow & \mathrm{Tor}_{i-1}(P, M) \longrightarrow \cdots
\end{array}$$

P is projective and thus flat, so the two modules in the left column are zero. Also, by inductive hypothesis, the two maps in the right two columns are isomorphic. But then that means $\mathrm{Tor}_i(M, N)$ and $\mathrm{Tor}_i(N, M)$ are the kernels of the two isomorphic maps from the third to the fourth column, so they are isomorphic. □

In particular, since $\mathrm{Tor}_i(M, \cdot)$ are the derived functors of the tensor product and thus form a universal δ-functor, and $\mathrm{Tor}_i(\cdot, M)$ are isomorphic to them in the way described above, $\mathrm{Tor}_i(\cdot, M)$ also form a universal δ-functor with

$$\mathrm{Tor}_i(\cdot, M) = L_i(\cdot \otimes_R M).$$

We'll now mention some more properties of flat modules, starting with a "locality" property:

Lemma 309

The following are equivalent:

1. M is flat over R,
2. $M_{\mathfrak{p}}$ is flat over $R_{\mathfrak{p}}$ for all prime ideals $\mathfrak{p}$,
3. $M_{\mathfrak{m}}$ is flat over $R_{\mathfrak{m}}$ for all maximal ideals $\mathfrak{m}$,

Proof. This is basically using the fact that localization is an exact functor. For (1) implies (2), suppose $0 \to N_1 \to N_2 \to N_3 \to 0$ is short exact, where each N_i is an $R_{\mathfrak{p}}$ module. Viewing them as R-modules, since M is flat, we get an exact sequence of R-modules

$$0 \to M \otimes_R N_1 \to M \otimes_R N_2 \to M \otimes_R N_3 \to 0.$$

But $M \otimes_R N_1$ is the same as $M \otimes_R (R_{\mathfrak{p}}) \otimes_{R_{\mathfrak{p}}} N_1)$, and now by the associative law (even when the tensor products are different) this is the same as $(M \otimes_R R_{\mathfrak{p}}) \otimes_{R_{\mathfrak{p}}} N_1 = M_{\mathfrak{p}} \otimes_{R_{\mathfrak{p}}} N_1$. So $M_{\mathfrak{p}}$ is indeed flat over $R_{\mathfrak{p}}$ because we get the short exact sequence we want.

(2) implies (3) is trivial because all maximal ideals are prime. Now for (3) implies (1), take an exact sequence of R-modules $0 \to N_1 \to N_2 \to N_3 \to 0$. We wish to tensor it with M, and because the tensor product is right exact we have a sequence

$$0 \to K \to M \otimes_R N_1 \to M \otimes_R N_2 \to M \otimes_R N_3 \to 0.$$

We wish to show $K = 0$, and as we have previously showed it is sufficient to show that $K_{\mathfrak{m}} = 0$ for all maximal ideals $\mathfrak{m}$. Since localization is exact, and $(M \otimes_R N)_{\mathfrak{m}} = M_{\mathfrak{m}} \otimes_{R_{\mathfrak{m}}} N_{\mathfrak{m}}$, localizing that sequence gives us

$$0 \to K_{\mathfrak{m}} \to M_{\mathfrak{m}} \otimes_{R_{\mathfrak{m}}} (N_1)_{\mathfrak{m}} \to M_{\mathfrak{m}} \otimes_{R_{\mathfrak{m}}} (N_2)_{\mathfrak{m}} \to M_{\mathfrak{m}} \otimes_{R_{\mathfrak{m}}} (N_3)_{\mathfrak{m}} \to 0.$$

But we also know that we have an exact sequence $0 \to (N_1)_{\mathfrak{m}} \to (N_2)_{\mathfrak{m}} \to (N_3)_{\mathfrak{m}} \to 0$ by localizing the original exact sequence, and then by assumption (3) that sequence is preserved under tensoring by $R_{\mathfrak{m}}$ so we must have $K_{\mathfrak{m}} = 0$, as desired. □

Lemma 310

If R is a noetherian **local ring** (meaning that it has a unique maximal ideal) and M is a finitely generated R-module, then M is free over R if and only if M is flat over R.

(So for noetherian rings, flatness is equivalent to being locally free.)

Proof. The forward direction is clear. For the backwards direction, we use Nakayama's lemma: looking at $M/(\mathfrak{m})M$ (where $\mathfrak{m}$ is the unique maximal ideal) we are finitely generated over the residue field and thus free. That means $M/(\mathfrak{m}M) \to (R/\mathfrak{m})^{\oplus d}$ is an isomorphism, and we can choose a surjection $R^{\oplus d} \to M$ which lifts that isomorphism by Nakayama. So there is a short exact sequence $0 \to K \to R^{\oplus d} \to M \to 0$ for some kernel K coming from that

surjection. Tensoring that by $(R/\mathfrak{m})\otimes_R$, we get

$$\mathrm{Tor}_1^R(R/\mathfrak{m}, M) \to K/\mathfrak{m} \to (R/\mathfrak{m})^{\oplus d} \xrightarrow{\cong} M/\mathfrak{m}M \to 0.$$

But now we can use the fact that M is flat again to see that the Tor module on the left is zero. Thus $K/\mathfrak{m} = 0$, but K is a submodule of a finitely-generated module over a noetherian ring, so it is also finitely generated. So by Nakayama's lemma again this means $K = 0$ and thus M is indeed free. □

Example 311

To calculate $\mathrm{Tor}_i^{\mathbb{Z}}(\mathbb{Z}/(n), \mathbb{Z}/(m))$, we take a projective resolution $\cdots \to 0 \to \mathbb{Z} \xrightarrow{\times m} \mathbb{Z} \to \mathbb{Z}/(m) \to 0$. To calculate Tor, we tensor with $\mathbb{Z}/(n)$ and take the $-i$th cohomology of $(\mathbb{Z}/(n) \xrightarrow{\times m} \mathbb{Z}/(n))$, where that map goes from degree -1 to 0.

Thus for $i = 0$ we want the cokernel of this map, which is $\mathbb{Z}/(m, n)$, and for $i > 1$ we just have zero. Finally for $i = 1$, we take the things in $\mathbb{Z}/(n)$ which are killed by multiplication by m, which is the set of $a + n\mathbb{Z} \in \mathbb{Z}$ such that n divides ma, which is the same as the a such that $\frac{n}{\gcd(n,m)}$ divides a. In other words, we end up finding that

$$\mathrm{Tor}_i^{\mathbb{Z}}(\mathbb{Z}/(n), \mathbb{Z}/(m)) = \begin{cases} \mathbb{Z}/(m, n) & i = 0, \\ \mathbb{Z}/(m, n) & i = 1, \\ 0 & i > 1. \end{cases}$$

though if we want to check functoriality with this we do need to be a bit careful with how we constructed the isomorphism $\frac{n}{\gcd(n,m)}\mathbb{Z}/(n\mathbb{Z}) \cong \mathbb{Z}/\gcd(n, m)\mathbb{Z}$ (it's not quite canonical).

So now if we consider the short exact sequence

$$0 \to \mathbb{Z}/(2) \xrightarrow{\times 2} \mathbb{Z}/(4) \to \mathbb{Z}/(2) \to 0$$

and we tensor with $\mathbb{Z}/(2)\otimes$, we end up with the long exact sequence

$$\cdots \to 0 \to \mathrm{Tor}_1^{\mathbb{Z}}(\mathbb{Z}/(2), \mathbb{Z}/(2)) \to \mathrm{Tor}_1^{\mathbb{Z}}(\mathbb{Z}/(2), \mathbb{Z}/(4)) \to \mathrm{Tor}_1^{\mathbb{Z}}(\mathbb{Z}/(2), \mathbb{Z}/(2)) \to \mathbb{Z}/(2) \xrightarrow{\times 2=0} \mathbb{Z}/(2) \xrightarrow{\cong} \mathbb{Z}/(2) \to 0,$$

where all three of the Tor terms are isomorphic to $\mathbb{Z}/(2)$. And then we see that the maps from left to right are an isomorphism (because it is injective), then zero (by exactness), then an isomorphism for the boundary map:

$$0 \to \mathbb{Z}/(2) \xrightarrow{\cong} \mathbb{Z}/(2) \xrightarrow{0} \mathbb{Z}/(2) \xrightarrow{\cong} \mathbb{Z}/(2) \xrightarrow{0} \mathbb{Z}/(2) \xrightarrow{\cong} \mathbb{Z}/(2) \to 0.$$

And in particular this means that the maps have shifted between the original sequence and the one between Tors.