Darek's notes on representation theory

(These notes should copy what I said in the lectures. They have been written hastily, so they may contain typos etc. But the math really should be OK.)

1 Modules

In its heart, representation theory is concerned with studying how groups can act on vector spaces. For that reason, the first structure we should introduce is a vector space that also "knows what to do" if elements of a group act on it. Of course, the group should respect the structure of the vector space too, so if $\boldsymbol{v}$ and $\boldsymbol{w}$ are two vectors in such a vector space and g is an element of the group G, we should have $g(\boldsymbol{v}+\boldsymbol{w}) = g\boldsymbol{v}+g\boldsymbol{w}$. Similarly, if α is a constant, we should have $g(\alpha\boldsymbol{v}) = \alpha g\boldsymbol{v}$.[1] Such a vector space is called a *G-module*.

The formal definition is as follows: a vector space V is called a *G-module* if for every $g \in G$ and $\boldsymbol{v} \in V$, a unique $g\boldsymbol{v} \in V$ is defined in such a way that: first, $g(\alpha\boldsymbol{v} + \beta\boldsymbol{w}) = \alpha g\boldsymbol{v} + \beta g\boldsymbol{w}$ (for all $v, w \in V$ and $\alpha, \beta \in \mathbb{C}$); second, if $g = g_1g_2$, then $g\boldsymbol{v} = g_1(g_2\boldsymbol{v})$ (for all $g_1, g_2 \in G$); third, $e\boldsymbol{v} = \boldsymbol{v}$ (where e is the identity element of the group).

As you can see, it's precise, but the essence is that "a *G-module* is a vector space on which elements of group G act as we would intuitively expect".

Here's an example: consider the group S_3 of permutations of three symbols. It has six elements: the identity permutation (), the three two-cycles (12), (23) and (31), and the two three-cycles (123) and (132). We take the vector space $\mathbb{R}^3$ with an orthonormal basis $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ and we say that each permutation of S_3 acts on this space by permuting the basis vectors accordingly. So (12) will switch $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ (geometrically, this is a reflection through some plane), (123) will permute the three vectors cyclically (which corresponds to the rotation by $2\pi/3$ around the axis $\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3$) etc. This is an S_3-module.

This module is a bit peculiar. It is easy to see that the vector $\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3$, or any multiple of it, is unaffected by any of the operations in S_3. This means that the one-dimensional subspace spanned by this vector is also an S_3-module in itself — its vectors still have the ability to be acted upon by elements of S_3. Moreover the result of any such operation will belong to the subspace, and it can never "leave" it. Hence this subspace is actually a *submodule* of the original module.

2 (Ir)reducibility of modules

Obviously, any module has two trivial submodules: the zero module (that contains only the zero vector and nothing else), and the module itself. These aren't very interesting. If a module as only these two submodules, it is called *irreducible*. On the other hand, if a module has a non-trivial submodule, we say it is *reducible*.

In our example of an S_3-module, we actually have a bit more: not only the subspace spanned by $\boldsymbol{e}_1+\boldsymbol{e}_2+\boldsymbol{e}_3$ is a submodule, but the "rest" of the vector space (a two-dimensional subspace spanned by $\boldsymbol{e}_1-\boldsymbol{e}_2$ and $\boldsymbol{e}_2-\boldsymbol{e}_3$) is a submodule as well. Hence we can write the module as a *direct sum*[2] of two smaller, but still non-trivial submodules. In this case, we say the module is *completely reducible*.

It may seem that this is just playing with words and that "reducibility" and "complete reducibility" must be the same thing. And often, it is. However, consider this example: we take the vector space $\mathbb{R}^2$, pick some basis in it, and we act on it with the group $\mathbb{Z}^3$ as follows: each integer k acts as a matrix $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$. Since $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & k+n \\ 0 & 1 \end{pmatrix}$, this vector space is a $\mathbb{Z}$-module. It is obviously reducible: the basis vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is unchanged by any of those matrices, so the subspace generated by it is a submodule. However, the

[1] This will probably feel more natural if we realize that the elements of groups often correspond to some geometrical operations like rotations, reflections or translations.

[2] A sum of vector spaces $V_1 + V_2$ is just a vector space formed by taking all possible linear combinations of vectors, one from V_1 and one from V_2. If the sum is *direct*, it just means that V_1 and V_2 are disjoint — they have just the zero vector in common.

[3] The group of all integers with addition.

"rest" of the vector space (the subspace generated by $\begin{pmatrix}0\\1\end{pmatrix}$) is not a submodule, because each group element makes these vectors "leak" into the $\begin{pmatrix}1\\0\end{pmatrix}$ direction. Hence this module is **NOT** completely reducible!

By the way: by now, you probably feel that this (ir)reducibility of modules must have something to do with the (ir)reducibility of representations. And it does. We use the modules because they offer a much clearer picture. Thanks to them, we see that the (ir)reducibility is the property of the module, i. e. of how the group acts on a vector space, and not of the representation matrices. The problem with matrices lies in the fact that we must always choose a basis in order to construct them; otherwise we wouldn't know which numbers to fill in. But that means that we are introducing extraneous information that obscures the problem and that we must actively grapple with. This only leads to unnatural definitions like "a bunch of matrices is completely reducible if we can change basis in such a way that they all simultaneously go to the block-diagonal form". Hence it is (in my opinion) better to use the language of modules.

3 Group algebra

Before we continue, we make a slight generalization. In a module, we can write things like $g_1\boldsymbol{v} + 2g_2\boldsymbol{v}$. This could be also understood as a formal sum of group elements, $g_1 + 2g_2$, acting on the vector $\boldsymbol{v}$. Of course, someone could object that groups contain no notion of addition and scalar multiplication, and so these sums do not make any sense. That is true to some extent.[4] However, once we have some vector space on which the group elements can act, we can easily act on it with such sums as well. In fact, we use sums of this type pretty often in physics — for instance, we often write things like $(\frac{\partial}{\partial x} - 3\frac{\partial}{\partial y})f$ and understand them to be the same as $\frac{\partial f}{\partial x} - 3\frac{\partial f}{\partial y}$. Here we use the same principle.

These formal sums can also be multiplied in an intuitive fashion. For instance, $(g_1 - g_2)(2g_1^{-1} + g_3) = 2 + g_1g_3 - 2g_2g_1^{-1} - g_2g_3$ — we assume the distributive law, numbers go to the front and the group elements multiply according to their multiplication table. All possible formal linear combinations of group elements, $\sum c_n g_n$, with the intuitive multiplication rules, make up an object called the *group algebra* of that group. As physicists, we assume that the coefficients of the linear combination can be any complex numbers, and we denote the group algebra of G simply by $\mathbb{C}G$.

Now we can also slightly extend our idea of a G-module: we can allow acting not only with the group elements, but with the formal linear combinations from $\mathbb{C}G$ as well. If we once more demand that these linear combinations do the intuitive thing, i. e. $(\sum c_n g_n)\boldsymbol{v} = \sum c_n(g_n\boldsymbol{v})$, we upgrade our G-module to a "$\mathbb{C}G$-module". Since this is so easy and natural to do, we will work with these "algebra-modules" from now on. It may seem like a useless formal trick, but it will come in handy later.

4 Maschke's theorem

Let's go back to the reducibility of modules now. We showed that reducibility and complete reducibility were different things. However, we can immediately forget about it, because there is an important result called *Maschke's theorem* which simply says that *for any finite group G, any $\mathbb{C}G$-module is always completely reducible*, so we always have the nice case.

What's more, the idea that makes it work is quite simple, and it can be gleaned from the S_3 example that we started with. That particular S_3-module decomposes to a direct sum of the one-dimensional submodule (a line in $\mathbb{R}^3$) and its orthogonal complement (the plane perpendicular to it). It turned out that the group action was unable to violate the orthogonality.

Hence we consider a $\mathbb{C}G$-module with a scalar product[5] $\langle \cdot, \cdot \rangle$, and we use it to build an *invariant scalar product* $\{\cdot, \cdot\}$ that does not change if we act with the same group element on both its operands, i. e. $\{g\boldsymbol{v}, g\boldsymbol{w}\} = \{\boldsymbol{v}, \boldsymbol{w}\}$ for any $g \in G$. If we can manage to do it, we win. If we find a submodule, we can use this

[4] This is why I say they are "formal" sums — they are objects obtained by writing some group elements with some numbers in front of them and sticking plus signs in between them, nothing more.

[5] Remember, the modules are just vector spaces, so they can have scalar product. If the module doesn't have it, we can always make one up by picking any basis and defining the scalar product by postulating that the basis vectors satisfy $\langle \boldsymbol{e}_k, \boldsymbol{e}_\ell \rangle = \delta_{k\ell}$.

invariant product to get its orthogonal complement. Moreover, if we take any vector $\boldsymbol{v}$ from the submodule and any $\boldsymbol{w}$ from the orthogonal complement, they will stay orthogonal no matter what group element is applied to them (since the product is invariant), and the vectors in the complem ent cannot leak into our submodule.

So we just have to find an invariant scalar product. We do it as follows:

$$\{\boldsymbol{v}, \boldsymbol{w}\} := \frac{1}{|G|} \sum_{g \in G} \langle g\boldsymbol{v}, g\boldsymbol{w} \rangle.$$

With this definition, $\{h\boldsymbol{v}, h\boldsymbol{w}\} = \frac{1}{|G|} \sum \langle gh\boldsymbol{v}, gh\boldsymbol{w} \rangle$. But for any group, each row and column of the multiplication table contains each group element exactly once. This means that if g goes through the whole group, gh does too, and the sum is unchanged, so we truly have $\{\boldsymbol{v}, \boldsymbol{w}\} = \{h\boldsymbol{v}, h\boldsymbol{w}\}$ for any $h \in G$. The product that we defined really is invariant, and Maschke's theorem is proved.

This is a result that will make our job much easier, because it can be applied repeatedly even to submodules of our module. Each time, we find out that reducible submodules (or their reducible submodules etc.) decompose into direct sums. This means that in the end, every $\mathbb{C}G$-module must decompose to a direct sum of irreducible submodules that do not interact with each other at all when group elements are applied. It turns out that for each group G, there is only limited number of different types of irreducible modules, and every $\mathbb{C}G$-module must be just a direct sum of some of these.

5 Homomorphisms

Now that we see that we are getting nice results for $\mathbb{C}G$-modules, we ask the question that mathematicians always ask when they encounter a new structure: Suppose that we have different objects equipped with the structure. What kind of maps can we make between them so that the structure is not ruined? Such maps always get the traditional name *homomorphism* (and we already saw one example of it, since we proved some properties of homomorphisms between groups).

Now suppose we have two $\mathbb{C}G$-modules, U and V, and some map $\varphi : U \to V$ between them. What properties should φ have in order to deserve the name of homomorphism between the $\mathbb{C}G$-modules? It must be a map that preserves their nice properties. So we must recollect what these properties are: firstly, the modules are *vector spaces*, secondly, they *know how* $\mathbb{C}G$ *acts on them* (and the elements of $\mathbb{C}G$ act in a certain consistent and intuitive manner).

A homomorphism between $\mathbb{C}G$-modules should preserve both of these things, so we define it as a *linear map* (i. e. it preserves the vector space structure) that also satisfies $\varphi(g\boldsymbol{u}) = g\varphi(\boldsymbol{u})$ (for any $\boldsymbol{u} \in U$ and any $g \in \mathbb{C}G$; of course, the action of g on $\varphi(\boldsymbol{u})$ on the right-hand side is that of the "destination" module V).[6]

Maps that satisfy the two conditions above (linear map + $\varphi(g\boldsymbol{u}) = g\varphi(\boldsymbol{u})$) should be called "$\mathbb{C}G$-module homomorphisms". But since that is quite a mouthful, and we will not consider homomorphisms of other objects than modules, we just shorten the name to a "$\mathbb{C}G$-homomorphism".

Now that we know what the $\mathbb{C}G$-homomorphisms are, we can start studying their properties. We know from linear algebra that each linear map defines two important subspaces of the "source" and "result" modules. It is the *kernel*, $\operatorname{Ker}\varphi$, which is the subspace of everything in U that gets sent to the zero in V, and the *image*, $\operatorname{Im}\varphi$, which is the subspace of all possible results that φ can give. We should have a closer look on what the group action does to these subspaces.

Let's have look at the kernel first. Suppose that some vector $\boldsymbol{k} \in U$ belongs to the kernel, which means that $\varphi(\boldsymbol{k}) = \boldsymbol{0}$. Now what happens if we act on $\boldsymbol{k}$ by some element $g \in \mathbb{C}G$? The map φ will send it to $\varphi(g\boldsymbol{k})$, but, since φ is a homomorphism, this must be the same thing as $g\varphi(\boldsymbol{k}) = g\boldsymbol{0} = \boldsymbol{0}$. So if $\boldsymbol{k}$ is in the kernel, every $g\boldsymbol{k}$ is there too. Hence $\operatorname{Ker}\varphi$ is a vector subspace of U (we know that from linear algebra) that also knows how $\mathbb{C}G$ acts on it (it's a subset of U and U knows it), and no matter what element of $\mathbb{C}G$ we apply, the result still belongs to the kernel. It cannot ever leak out of it. Hence *the kernel of any* $\mathbb{C}G$*-homomorphism is a* ***submodule*** *of* U.

[6] In words, the second condition means that any element $g \in \mathbb{C}G$ acts on the "source" module U in the same way as it acts on the "result" module V, which is what we need.

What about the image? Let's say that some vector $\boldsymbol{v} \in V$ belongs to the image. This means that it can be obtained as a result of applying the homomorphism φ on some $\boldsymbol{u} \in U$, so $\boldsymbol{v} = \varphi(\boldsymbol{u})$. However, if we apply φ on $g\boldsymbol{u}$ (for any $g \in \mathbb{C}G$), we get $\varphi(g\boldsymbol{u}) = g\varphi(\boldsymbol{u}) = g\boldsymbol{v}$, because φ is a $\mathbb{C}G$-homomorphism. So every $g\boldsymbol{v}$ is in the image as well, and *the image of any $\mathbb{C}G$-homomorphism is a submodule of V.*

6 Schur's lemma

Perhaps you're now asking "fine, $\operatorname{Ker}\varphi$ and $\operatorname{Im}\varphi$ are submodules. So what?" But we really should be interested in that, since we want to study irreducible modules (i. e. modules without submodules), and these simple facts already put stringent conditions on maps between such modules.

So suppose that we have two *irreducible* $\mathbb{C}G$-modules, U and V, and a $\mathbb{C}G$-homomorphism φ between them. What does φ look like? There's not really much freedom here, because $\operatorname{Ker}\varphi$ must be a submodule of U, and since U is irreducible, it has only the two trivial submodules: the zero and U itself. In the first case, φ must be a bijection[7], and the image is all of V. In the second case, φ maps everything to zero, and the image is the zero module. Nothing else is possible.

This gives the following: *Any $\mathbb{C}G$-homomorphism between two irreducible $\mathbb{C}G$-modules is either a bijection, or a zero map.* This result is called the *Schur's lemma* and though it is simple, it is powerful enough to serve as a base out of which everything in the representation theory may be derived.

There is also a related question: what $\mathbb{C}G$-homomorphisms can exist from an irreducible module U back to itself? Such a linear map $\varphi : U \to U$ can be represented by a square matrix M (if we pick some basis in U). This matrix has some eigenvalues given by the equation $\det(M - \lambda E) = 0$. We pick one of them and consider the map $\varphi - \lambda$. It is also a $\mathbb{C}G$-homomorphism from U to U, so Schur's lemma may be applied to it. Hence $\varphi - \lambda$ must be a bijection, or zero. But its determinant is zero (λ was picked in such a way that it would be!), so it cannot be a bijection. Hence $\varphi - \lambda = 0$ and *the only possible $\mathbb{C}G$-homomorphisms from an irreducible module to itself is a multiplication by a constant.*[8] This result is often also called the Schur's lemma and it too will be very useful for us in the following.

[7] Each $\boldsymbol{u} \in U$ is mapped to something in V, and each to something different. If two $\boldsymbol{u}_1 \neq \boldsymbol{u}_2$ map to the same $\in V$, their difference must map to zero due to linearity, and so it must be in the kernel. But the kernel has only the zero vector in it.

[8] Here it is crucial that everything we do is over complex numbers. Then $\det(M - \lambda E)$ is a polynomial in λ with complex coefficients, so it always must have a complex root λ. In real numbers, for instance, this would not work (for instance, the real matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ does not have a real eigenvalue)! However, as physicists, we love complex numbers and we use them everywhere, so we don't need to worry about this result being true only when complex numbers are allowed. (In fact there are other "algebraically closed fields" that have this property. But we should leave them to the mathematicians.)