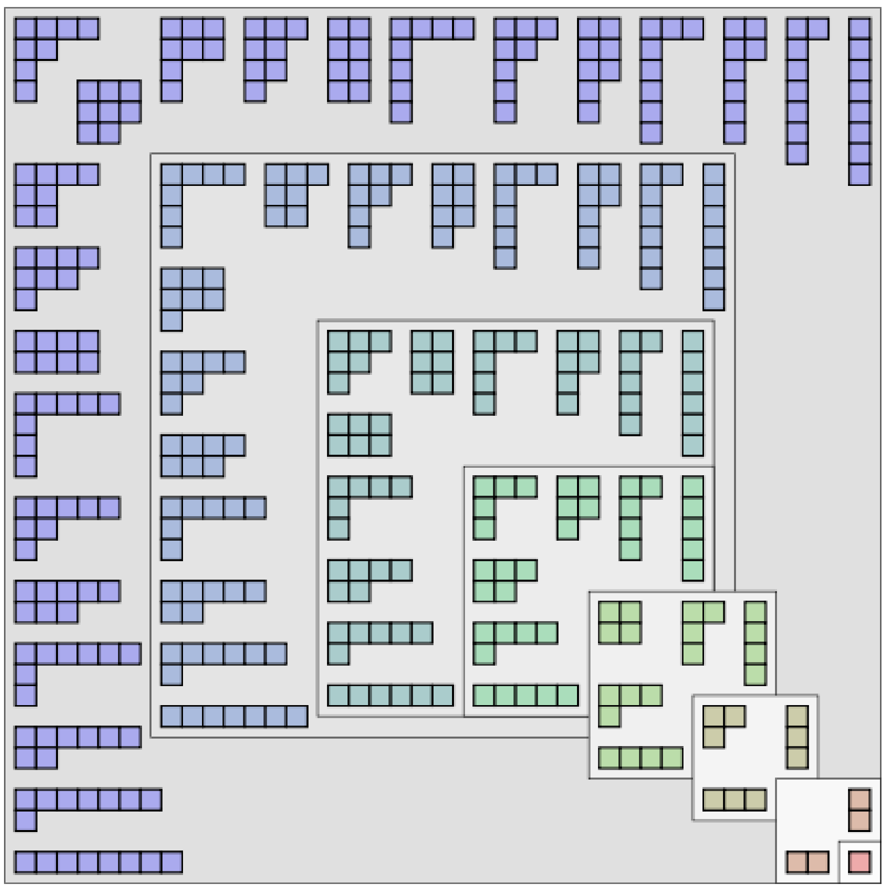
Young Diagrams and Classical Groups

John Baez
Department of Mathematics
University of California, Riverside

September 28, 2022

based on "week157" of *This Week's Finds*

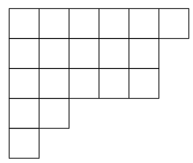


Mathematics and physics rely a lot on *symmetry* to simplify problems, and there are two kinds of diagrams that show up a lot in this context: Dynkin diagrams and Young diagrams. Dynkin diagrams first show up when you study shapes with lots of reflection symmetries, like crystals and Platonic solids. They wind up being good for all sorts of other stuff, like classifying simple Lie groups and their representations. But what about Young diagrams? These are also important for studying group representations, but for a more limited class of groups: the “classical” groups. Representations of classical groups are used a lot in quantum physics, from particle physics through nuclear physics and atomic physics up to chemistry. So Young diagrams are not only beautiful, they’re practical.

My goal is to explain how Young diagrams are used to classify representations of classical groups. I won’t prove much, just sketch the ideas. First I’ll explain classical groups and group representations. But even before that, I should say what’s a Young diagram.

Young diagrams

Here is an example of a Young diagram:

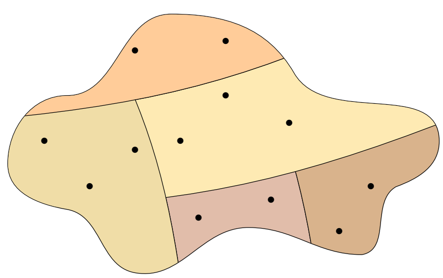


All the information here is captured by the number of boxes in each row:

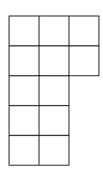
$$6 \geq 5 \geq 5 \geq 2 \geq 1$$

So, we can define a **Young diagram** to be a finite sequence of natural numbers $n_1 \geq n_2 \geq \cdots \geq n_k > 0$. We say k is the number of **rows** and n_1 is the number of **columns**. We say n_i is the number of **boxes** in the ith column, and $n = \sum_i n_i$ is the total number of boxes.

Young diagrams with n boxes classify partitions of an n-element set, up to isomorphism. For example, this partition:

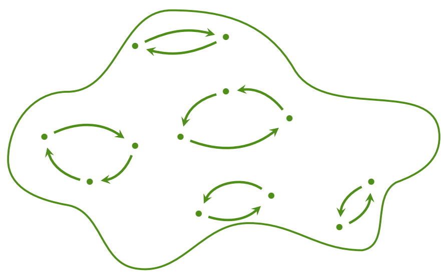


gives this Young diagram, whose rows list how many points are in each part:



But the Young diagram does not record which point of our set lies in which part, so Young diagrams classify partitions only "up to isomorphism".

Young diagrams with n boxes also classify permutations of an n-element set up to isomorphism. For example this permutation:



gives the same Young diagram we have just seen. But any isomorphic permutation would give the same Young diagram.

What's an "isomorphic permutation", exactly? Let's look at an example. Permutations of the set $\{1, \ldots, n\}$ form the **symmetric group** S_n. Say we have any permutation $g \in S_n$, like this:

$$\begin{aligned} 1 &\to 2 \\ 2 &\to 4 \\ 3 &\to 3 \\ 4 &\to 1 \\ 5 &\to 6 \\ 6 &\to 5 \\ 7 &\to 7 \end{aligned}$$

Note that 1 gets mapped to 2, which gets mapped to 4, which gets mapped back to 1 again. Similarly, 5 gets mapped to 6, which gets mapped back to 5. The number 3 gets mapped to itself right away, as does 7. No matter where we start, we always cycle back eventually. So our permutation consists of a bunch of cycles:

$$(1, 2, 4)(5, 6)(3)(7)$$

and this "cycle decomposition" completely describes the permutation. To simplify life, we always write down these cycles in order of decreasing length. We also write the lowest number in each cycle first.

Now suppose we conjugate our permutation g by some other permutation, say h. This gives the permutation hgh^{-1}. How does the cycle decomposition of this compare with that of g? It looks very similar! For example, it might look like this:

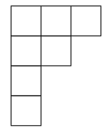
$$(2, 7, 6)(1, 3)(4)(5)$$

There are the same number of cycles, each the same length as before. The only thing that changes are the numbers in each cycle. These get switched around by means of the permutation h.

In short, when we conjugate a permutation, all that remains unchanged is the picture we get by writing down its cycle decomposition and blotting out the specific numbers in each cycle, like this:

$$(\square, \square, \square)(\square, \square)(\square)(\square)$$

If we write each cycle as a row of boxes, we get a Young diagram:



Classical groups, and a classical monoid

Now, what are the classical groups? As with composers of music, there's no precise list of groups that count as "classical". But in general, a classical group should consist of linear transformations that preserve some nice geometrical structure on a vector space. Some good examples are:

- The **general linear group** $\mathrm{GL}(N, \mathbb{C})$, consisting of all invertible linear transformations of $\mathbb{C}^N$, or in other words, all $N \times N$ complex matrices with nonzero determinant.
- The **special linear group** $\mathrm{SL}(N, \mathbb{C})$, consisting of all linear transformations of $\mathbb{C}^N$ with determinant 1.
- The **unitary group** $\mathrm{U}(N)$, consisting of all unitary linear transformations of $\mathbb{C}^N$.
- The **special unitary group** $\mathrm{SU}(N)$, consisting of all unitary linear transformations of $\mathbb{C}^N$ with determinant 1.

These are the Bach, Haydn, Mozart and Beethoven of classical groups. Representations of all four can be classified with the help of Young diagrams.

We may also consider this an honorary classical group, even though it's defined in terms of a *set* rather than a *vector space*:

- The symmetric group S_n, consisting of all permutations of the set $\{1, \ldots, n\}$.

Representations of this group are also classified using Young diagrams—and as we'll see, S_n plays a starring role in the whole story.

There's another key actor whose representations are classified by Young diagrams. It deserves to be called a "classical monoid":

- The **full linear monoid** $\mathrm{End}(\mathbb{C}^n)$, consisting of all linear transformations of $\mathbb{C}^N$, or in other words, all $N \times N$ matrices.

A **monoid** is a set with an associative multiplication and identity, but not necessarily inverses. Here I am making $\mathrm{End}(\mathbb{C}^n)$ into a monoid where the multiplication is composition of transformations—or in low-brow terms, matrix multiplication. This monoid is so classical that people don't even call it that! Perhaps the common prejudice in favor of groups and against other monoids is to blame. As we'll see, the full linear monoid is a bit like the composer Palestrina, who is not considered a classical composer, yet who set the stage for the music we call classical.

Representations

Groups feel sad unless they are acting as symmetries of something. Monoids feel the same way—or even worse, because they're less loved than groups. This why we should study representations of groups and monoids. A **homomorphism** of monoids, say $\rho\colon M \to N$, is a function with

$$\rho(mm') = \rho(m)\rho(m') \text{ for all } m, m' \in M \text{ and } \rho(1) = 1.$$

A **representation** of a monoid M on a vector space V is a homomorphism

$$\rho\colon M \to \mathrm{End}(V)$$

where $\mathrm{End}(V)$ consists of all linear transformations of V, made into a monoid using composition. A representation lets us take an element $m \in M$ and make it act on a vector $v \in V$ to get a new vector $\rho(m)v$, in such a way that

$$\rho(mm')v = \rho(m)\rho(m')v \text{ and } \rho(1)v = v.$$

So now our monoid is doing something, not just sitting there moping!

But a representation is still lonely in isolation. To solve this problem we define morphisms between representations of given monoid, getting an entire *category* of representations. Given two representations $\rho\colon M \to \mathrm{End}(V), \sigma\colon M \to \mathrm{End}(W)$, a **morphism** from the first to the second is a linear map $f\colon V \to W$ such that

$$f(\rho(m)v) = \sigma(m)f(v)$$

for all $v \in V$. That is: acting and then mapping is the same as mapping and then acting. Thanks to how f slips from outside to inside in this equation, morphisms of representations are also called **intertwining operators**.

An **isomorphism** is just a morphism with an inverse, and an isomorphism of representations is also commonly called an **equivalence**. We won't do much with categories here except for classifying representations "up to isomorphism": when we do that, we don't distinguish between isomorphic representations. But studying the whole category of representations of a monoid, all at once, is a good way to get deeper insights in representation theory.

The simplest representations are those on finite-dimensional vector spaces—so henceforth:

We assume all vector spaces under discussion are finite-dimensional, without even mentioning it!

And instead of trying to study *all* finite-dimensional representations, I will focus on the "irreducible" ones, which serve as building blocks for more complicated ones. For example in particle physics we use irreducible representations to describe elementary particles. A representation ρ of a monoid on a vector space V is **irreducible** if V has no subspaces invariant under all the transformations $\rho(g)$, except for $\{0\}$ and V itself. "Irreducible representations" is a bit of a mouthful, so we also call them **irreps** for short.

Why are irreducible representations important? Arguably the "indecomposable" representations are even more important to us here. Given two representations of a monoid, say $\rho\colon M \to \mathrm{End}(V)$ and $\rho'\colon M \to \mathrm{End}(V')$, there is a representation on $V \oplus V'$ called their **direct sum**:

$$\rho \oplus \rho'\colon M \to \mathrm{End}(V \oplus V')$$

given by

$$(\rho \oplus \rho')(m)(v, v') = (\rho(m)v, \rho'(m)v').$$

A representation is **indecomposable** if it is not isomorphic to a direct sum of representations except for the 0-dimensional representation and itself. Using an inductive argument we can show that every representation is a direct sum of indecomposable representations. That is, we can break apart any representation into smaller pieces until we reach pieces that can't be broken apart any further.

It is easy to see that any irreducible representation is indecomposable. The converse is not always true. However, for all the monoids we shall consider here, and the kinds of representations we consider here, indecomposability is equivalent to irreducibility! And since "irrep" is such a handy word, we shall talk about irreducibility rather than indecomposability.

S_n

Amazingly, Young diagrams can be used to classify the irreps, or at least the "nice" ones, of all five classical groups I listed—$\mathrm{GL}(N, \mathbb{C})$, $\mathrm{SL}(N, \mathbb{C})$, $\mathrm{U}(N)$, $\mathrm{SU}(N)$ and S_n—together with the classical monoid $\mathrm{End}(\mathbb{C}^N)$. Let me sketch how this goes. We'll start with the symmetric groups S_n, which are the most important of all.

Remember, I've shown how conjugacy classes of permutations in S_n correspond to Young diagrams with n boxes. Now I want to do the same for irreducible representations of S_n. This is cool for the following reason: for any finite group, the number of irreducible representations is the same as the number of conjugacy classes of group elements! But in general there's no natural one-to-one correspondence between irreducible representations with conjugacy classes. The group S_n just happens to be specially nice in this way.

To get started I should tell you some stuff that work for any finite group. Suppose G is a finite group. Then G has only finitely many irreps, all finite-dimensional. Every finite-dimensional representation of G is a direct sum of copies of these irreps.

To get our hands on these irreps, let $\mathbb{C}[G]$ be the space of formal linear combinations of elements of G. This is called the **group algebra** of G, since it becomes an algebra using the product in G. With some work, one can show that $\mathbb{C}[G]$ is isomorphic to an algebra of block diagonal matrices. For example, $\mathbb{C}[S_3]$ is isomorphic to the algebra of matrices of this form:

$$\begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}$$

where the $*$ entries can be any complex number whatsoever. Since matrices act on vectors by matrix multiplication, we can use this to get a bunch of representations of $\mathbb{C}[G]$, and thus of G — one representation for each block. And this trick gives us all the irreps of G! For example, S_3 has two 1-dimensional irreps, coming from the two 1×1 blocks in the above matrix, and one 2-dimensional irrep, coming from the 2×2 block.

In fact, we can actually concoct these irreps as subspaces of $\mathbb{C}[G]$. One way is to find elements of $\mathbb{C}[G]$ with a single 1 on the diagonal of one block and zero everywhere else, like these:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{p_1} \quad \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{p_2} \quad \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{p_3}$$

If we can find these guys, right multiplying by them will project down to various subspaces of $\mathbb{C}[G]$, namely

$$\{ap_i \mid a \in \mathbb{C}[G]\}.$$

And these subspaces will be irreps of G, as you can check using our description of $\mathbb{C}[G]$ as an algebra of block diagonal matrices.

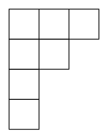
How do we find these guys p_i in $\mathbb{C}[G]$? That takes work! But for starters, notice that:

- They are **idempotent**: $p_i^2 = p_i$.
- They are **minimal**: if p_i is the sum of two idempotents, one of them must be zero.
- They are **separated**: if $i \neq j$ we have $p_i a p_j = 0$ for all $a \in \mathbb{C}[G]$.

Indeed they form a large-as-possible collection of separated minimal idempotents: as many as the number of irreps G—or equivalently, the number of conjugacy classes in G.

To go further, we need to know more about our group G. So now I'll take G to be S_n and tell you how to get separated minimal idempotents. We'll get one for each Young diagram with n boxes! Since there's as many conjugacy classes in S_n as n-box Young diagrams, that will mean we've got a large-as-possible collection.

Here's how it works. Say we have a Young diagram with n boxes, like this:



Then we can pack it with numbers from 1 to n like this:

$$\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & 5 & \multicolumn{1}{c}{} \\ \cline{1-2} 6 & \multicolumn{2}{c}{} \\ \cline{1-1} 7 & \multicolumn{2}{c}{} \\ \cline{1-1} \end{array}$$

There are a bunch of permutations in S_n called **row permutations** that only permute the numbers within each row of our Young diagram. And there are a bunch called **column permutations** that only permute the numbers within each column.

We can form an idempotent p in $\mathbb{C}[S_n]$ that symmetrizes over all row permutations. We get p by taking the sum of all row permutations divided by the number of row permutations:

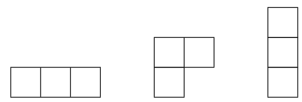
$$p = \frac{1}{|R|} \sum_{\sigma \in R} \sigma \in \mathbb{C}[S_n]$$

where R is the set of row permutations. Similarly, we can form an idempotent q in $\mathbb{C}[S_n]$ that antisymmetrizes over all column permutations. We get q by taking the sum of all *even* column permutations minus the sum of all *odd* column permutations, and then dividing by the total number of column permutations:

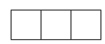
$$q = \frac{1}{|C|} \sum_{\sigma \in C} \mathrm{sgn}(\sigma)\sigma \in \mathbb{C}[S_n]$$

where C is the set of column permutations. Now here's the cool part: up to a constant factor, pq is a minimal idempotent in $\mathbb{C}[S_n]$! Even better, this procedure gives exactly one minimal idempotent for each block in the block matrix description of $\mathbb{C}[S_n]$. This isn't obvious at all—it takes real work to prove—but it's the crucial fact that connects n-box Young diagrams to representations of S_n.

Consider $n = 3$, for example. There are 3 Young diagrams in this case:



so S_3 has 3 irreps, confirming something I already said. For the long squat diagram

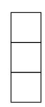


the column permutations are trivial, so the minimal central idempotent is just p. That is, it just "symmetrizes": it's the sum of all 3! permutations in S_3, divided by 3!. It winds up giving a 1×1 block in

$$\mathbb{C}[S_3] \cong \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}$$

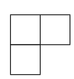
and thus a 1-dimensional representation of S_3. This is the **trivial representation** where every element of S_3 acts as the identity operator on $\mathbb{C}$. Every monoid has a trivial representation.

For the tall skinny diagram



the row permutations are trivial, so the minimal idempotent is just q. That is, it just "antisymmetrizes": it's the sum of all 3! permutations times their signs, divided by 3!. This gives the other 1-dimensional representation of S_3: the **sign representation** where each permutation acts on $\mathbb{C}$ as multiplication by its sign.

The remaining 3-box Young diagram



is a bit trickier. It gives a minimal idempotent that does a more interesting mix of row symmetrization and column antisymmetrization. This gives the 2-dimensional representation of S_3.

Here's a more concrete way to describe this representation. You can think of S_3 as the symmetries of an equilateral triangle. If you draw such a triangle in the plane, centered at the origin, each symmetry of this triangle gives a linear transformation of $\mathbb{R}^2$, or in other words a 2×2 real matrix. But you can think of this as a complex 2×2 matrix! This trick defines a homomorphism $\rho\colon S_3 \to \mathrm{End}(\mathbb{C}^2)$, and this is our representation.

$\mathrm{End}(\mathbb{C}^N)$

We could go on thinking about Young diagrams and representations of the symmetric groups S_n for a long time. People have spent their lives on this! But before we get too old, let's see how Young diagrams give representations of the four other classical groups.

It's actually best to start with the full linear monoid $\mathrm{End}(\mathbb{C}^N)$, since those four classical groups are all contained in this. Indeed we have monoid homomorphisms like this, all given by inclusions:

$$\begin{tikzcd}
\mathrm{SU}(N) \arrow[r] \arrow[d] & \mathrm{U}(N) \arrow[d] & \\
\mathrm{SL}(N,\mathbb{C}) \arrow[r] & \mathrm{GL}(N,\mathbb{C}) \arrow[dr] & \\
& & \mathrm{End}(\mathbb{C}^N)
\end{tikzcd}$$

Whenever you have a monoid homorphism $f\colon M \to M'$ and a representation of M', say $\rho\colon M' \to \mathrm{End}(V)$, you can compose them and get a representation of M. So, representations of $\mathrm{End}(\mathbb{C}^N)$ give representations of all four classical groups I listed—and this is actually how we'll get our hands on irreps of these classical groups.

So let's try to understand representations of the monoid $\mathrm{End}(\mathbb{C}^N)$. For starters, it has an representation on $\mathbb{C}^N$ called the **tautologous representation**, where each transformation acts on vectors in $\mathbb{C}^N$ in the obvious way. In other words, this representation is the identity homomorphism $1\colon \mathrm{End}(\mathbb{C}^N) \to \mathrm{End}(\mathbb{C}^N)$. This is actually an irrep.

How can we get other irreps of $\mathrm{End}(\mathbb{C}^N)$? One way to get new representations from old is by tensoring them. If we have two representations $\rho\colon M \to \mathrm{End}(V)$, $\rho'\colon M \to \mathrm{End}(V')$ of any monoid, we get a new one called $\rho \otimes \rho'$ with

$$\begin{array}{rcl}
\rho \otimes \rho'\colon M & \to & \mathrm{End}(V) \otimes \mathrm{End}(V') \cong \mathrm{End}(V \otimes V') \\
m & \mapsto & \rho(m) \otimes \rho(m').
\end{array}$$

So, one thing we can do is take the tautologous representation of $\mathrm{End}(\mathbb{C}^N)$ and tensor it with itself a bunch of times, say n times, getting a representation on

$$\underbrace{\mathbb{C}^N \otimes \mathbb{C}^N \otimes \ldots \otimes \mathbb{C}^N}_{n \text{ copies}}$$

There's no reason in the world this new representation should be irreducible. But we can try to chop it up into irreducible bits. And the easiest way is to look for bits that transform in nice ways when we permute the n copies of $\mathbb{C}^N$. In physics lingo, we have a space of tensors with n indices, and we can look for subspaces consisting of tensors that transform in specified ways when we permute the indices. For example, there will be a subspace consisting of "totally symmetric" tensors that don't change at all when we permute the indices, and a subspace of "totally antisymmetric" tensors that change sign whenever we interchange two indices, and so on.

But to make the "and so on" precise, we need Young diagrams! After all, these describe all the representations of the permutation group.

Here's how it works. The space

$$(\mathbb{C}^N)^{\otimes n} = \underbrace{\mathbb{C}^N \otimes \mathbb{C}^N \otimes \cdots \otimes \mathbb{C}^N}_{n \text{ copies}}$$

is not only a representation of $\mathrm{End}(\mathbb{C}^N)$; it's also a representation of $\mathbb{C}[S_n]$, coming from permutations of the n factors. And the actions of these two monoids commute! This means that we can chop up $(\mathbb{C}^N)^{\otimes n}$ into subspaces using the minimal idempotents in S_n that we get from Young diagram, and each of these subspaces will be a representation of $\mathrm{End}(\mathbb{C}^N)$.

That much should be obvious. The really cool part is that all these subspaces are *irreducible* representations of $\mathrm{End}(\mathbb{C}^N)$. This is much less obvious! The reason, ultimately, is that the linear transformations of $(\mathbb{C}^N)^{\otimes n}$ that commute with all transformations coming from the representation of $\mathrm{End}(\mathbb{C}^N)$ on this space are precisely those coming from $\mathbb{C}[S_n]$. This is half of a result called "Schur–Weyl duality". And I can't resist mentiong the other half, though we don't need it here. It says that the linear transformations of $(\mathbb{C}^N)^{\otimes n}$ that commute with all transformations coming from the representation of $\mathbb{C}[S_n]$ on this space are precisely those coming from $\mathrm{End}(\mathbb{C}^N)$.

As you can see, there is some serious math going on here. In any event, each Young diagram gives an irrep of $\mathrm{End}(\mathbb{C}^N)$. Let's see how this works in a few examples.

If we take $n = 3$, then S_3 acts on

$$(\mathbb{C}^N)^{\otimes 3} = \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$$

So, we get some irreps of $\mathrm{End}(\mathbb{C}^N)$ from 3-box Young diagrams. As we've seen, the long squat Young diagram

$$\begin{array}{|c|c|c|}\hline & & \\ \hline \end{array}$$

gives the minimal idempotent that just "symmetrizes". So it gives an irrep of $\mathrm{End}(\mathbb{C}^N)$ on the space of **symmetric tensors of rank 3**:

$$S^3(\mathbb{C}^N) = \Big\langle \frac{1}{3!} \sum_{\sigma \in S_n} v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes v_{\sigma(3)} \;\Big|\; v_1, v_2, v_3 \in \mathbb{C}^N \Big\rangle$$

where the angle brackets mean we take all linear combinations of vectors of this form. Similarly, the tall skinny Young diagram

$$\begin{array}{|c|}\hline \\ \hline \\ \hline \\ \hline \end{array}$$

gives the minimal idempotent that "antisymmetrizes". So it gives an irrep of $\mathrm{End}(\mathbb{C}^N)$ on the space of **antisymmetric tensors of rank 3**:

$$\Lambda^3(\mathbb{C}^N) = \Big\langle \frac{1}{3!} \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma)\, v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes v_{\sigma(3)} \;\Big|\; v_1, v_2, v_3 \in \mathbb{C}^N \Big\rangle.$$

All this works the same way for any other number replacing 3. The other 3-box Young diagram

$$\begin{array}{|c|c|}\hline & \\ \hline \\ \cline{1-1} \end{array}$$

is more tricky. To get its minimal idempotent up to a constant factor, you need to first antisymmetrize over column permutations of the numbers here:

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1} \end{array}$$

and then symmetrize over row permutations. Then you apply the resulting element of $\mathbb{C}[S_3]$ to all vectors $v_1 \otimes v_2 \otimes v_3$, and take all linear combinations of what you get. I could write down the formulas, but you probably wouldn't enjoy it. In math, some things are more fun to do than to watch.

When you think about this game works, you'll notice that some of irreps we get are a bit silly. If we have a Young diagram with more than N rows, we'll be antisymmetrizing over a tensor product of more than N vectors in $\mathbb{C}^N$, which always gives zero. So such Young diagrams give zero-dimensional representations of $\mathrm{End}(\mathbb{C}^N)$. We can ignore these. Indeed, most people decree that zero-dimensional representations don't even count as irreducible, just as the number 1 isn't prime. Let's do that from now on.

With this convention in place, we get an irrep of $\mathrm{End}(\mathbb{C}^N)$ from each Young diagram with at most N rows. And they're all different: that is, distinct Young diagrams with at most N rows give nonisomorphic representations.

Do we get *all* the irreps of $\mathrm{End}(\mathbb{C}^N)$ from Young diagrams with at most N rows? No, alas. Suppose we have a representation ρ of $\mathrm{End}(\mathbb{C}^N)$ that arises from a Young diagram. Say it acts on some vector space L. If we pick a basis for L, we can write each linear transformation $\rho(x)\colon L \to L$ as a matrix, and you can check that the matrix entries of $\rho(x)$ are polynomials in the entries of the original matrix $x \in \mathrm{End}(\mathbb{C}^N)$. Thus we say ρ is a **polynomial representation**—and we see that Young diagrams can only give us polynomial representations of $\mathrm{End}(\mathbb{C}^N)$.

Thus, as soon as you find a irrep of $\mathrm{End}(\mathbb{C}^N)$ that's not a polynomial representation, you'll know that you can't get *all* the irreps of $\mathrm{End}(\mathbb{C}^N)$ from Young diagrams. And such an irrep is not hard to find. For example, consider the representation

$$\begin{array}{cccc} \rho\colon & \mathrm{End}(\mathbb{C}^N) & \to & \mathrm{End}(\mathbb{C}^N) \\ & T & \mapsto & \overline{T} \end{array}$$

that takes the complex conjugate of each entry of an $N \times N$ matrix. There are many more.

But the next best thing is true: every polynomial irrep of $\mathrm{End}(\mathbb{C}^N)$ comes from a Young diagram. In fact there is a one-to-one correspondence between these things:

- polynomial irreps of $\mathrm{End}(\mathbb{C}^n)$, up to isomorphism
- Young diagrams with $\leq N$ rows.

Thus, we say that Young diagrams with at most N rows *classify* polynomial irreps of $\mathrm{End}(\mathbb{C}^N)$. This remarkable fact is the basic link between Young diagrams and representations of the classical groups. Let's see how to use it.

$\mathrm{GL}(N, \mathbb{C})$

Let's start with the biggest of the classical groups, the general linear group $\mathrm{GL}(N, \mathbb{C})$. Consider its inclusion in $\mathrm{End}(\mathbb{C}^N)$:

$$\mathrm{GL}(N, \mathbb{C}) \to \mathrm{End}(\mathbb{C}^N)$$

Composing this with any polynomial irrep of $\mathrm{End}(\mathbb{C}^N)$, we get a representation of $\mathrm{GL}(N, \mathbb{C})$. In fact it is an irrep. We don't get all the irreps of $\mathrm{GL}(N, \mathbb{C})$, but we get all the polynomial irreps: that is, those whose matrix entries are polynomials in the matrix entries of the element $g \in \mathrm{GL}(N, \mathbb{C})$ they depend on.

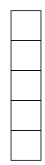
Furthermore, since $\mathrm{GL}(N,\mathbb{C})$ is dense in $\mathrm{End}(\mathbb{C}^n$ and polymomials are continuous, distinct polynomial irreps of $\mathrm{End}(\mathbb{C}^N)$ give distinct polynomial irreps of $\mathrm{GL}(N,\mathbb{C})$. Even better, every polynomial irrep arises from one of $\mathrm{End}(\mathbb{C}^n)$. Using these ideas and our previous results on representations of $\mathrm{End}(\mathbb{C}^N)$, we can show that there is a one-to-one correspondence between these things:

- polynomial irreps of $\mathrm{GL}(N,\mathbb{C})$, up to isomorphism
- Young diagrams with $\leq N$ rows.

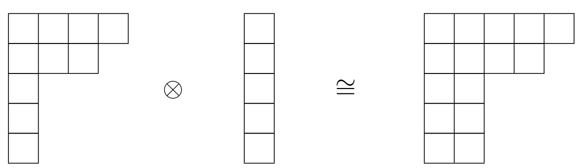
Even better, every polynomial representation of $\mathrm{GL}(N,\mathbb{C})$ can be written as a direct sum of polynomial irreps.

However, there are plenty of non-polynomial irreps of $\mathrm{GL}(N,\mathbb{C})$: not only those coming from the non-polynomial irreps of $\mathrm{End}(\mathbb{C}^N)$, but also others. The reason is that a matrix in $\mathrm{GL}(N,\mathbb{C})$ has nonzero determinant, so we can cook up representations involving the inverse of the determinant, which is not a polynomial.

The 1-dimensional irrep of $\mathrm{GL}(N,\mathbb{C})$ sending each matrix g to $\det(g)$, called the **determinant representation**. This is a polynomial irrep, so it must come from a Young diagram. Indeed it comes from tall skinny Young diagram with one column and N rows, e.g.



when $N = 5$. If we have any irrep of $\mathrm{GL}(N,\mathbb{C})$ coming from a Young diagram, tensoring it with the determinant representation gives a new irrep described by a Young diagram with an extra column with N rows, like this:



However, there's also a 1-dimensional irrep of $\mathrm{GL}(N,\mathbb{C})$ that sends $g \in \mathrm{GL}(N,\mathbb{C})$ to $\det(g)^{-1}$. This is called the **inverse** of the determinant representation, both for the obvious reason and because when you tensor it with the determinant representation you get the trivial representation. Since $\det(g)^{-1}$ is not a polynomial in the matrix entries of g, this not a polynomial representation. But it is still an **algebraic representation**: one whose matrix entries are rational functions of the matrix entries of g.

Algebraic representations are the kind most natural in algebraic geometry. Indeed $\mathrm{GL}(N,\mathbb{C})$ is a **linear algebraic group** over $\mathbb{C}$: that is, a group in the category of affine algebraic varieties over the complex numbers. When people talk about representations of linear algebraic groups, they usually mean algebraic representations.

So, fans of algebraic geometry will be glad to know that algebraic irreps of $\mathrm{GL}(N,\mathbb{C})$ can all be built by taking a polynomial irrep and tensoring it with the inverse of the determinant representation some number of times. This in turn means we can describe any algebraic irrep of $\mathrm{GL}(N,\mathbb{C})$ using a Young diagram with *fewer than* N rows together with an integer k. The Young diagram gives a representation ρ, and then we form the representation on the same space where g acts by $\det(g)^k \rho(g)$. If $k \geq 0$ this is the same as tacking on k extra columns with N rows to our Young diagram, but the procedure also makes sense for $k < 0$. We get a one-to-one correspondence between these things:

- algebraic irreps of $\mathrm{GL}(N, \mathbb{C})$, up to isomorphism
- pairs consisting of a Young diagram with $< N$ rows and an integer.

If you like, you can think of such a pair as a funny sort of Young diagram with $\leq N$ rows where the number of columns with N rows can be any integer—even a negative number!

This is the story for irreps, but what about more general representations? It's as nice as it could be: every algebraic representation of $\mathrm{GL}(N, \mathbb{C})$ is a direct sum of algebraic irreps.

If you don't yet love algebraic geometry, you may prefer to think of $\mathrm{GL}(N, \mathbb{C})$ as a **complex Lie group**: a group in the category of complex manifolds. When we talk about a representation of a complex Lie group G, we usually mean an **complex-analytic representation**: a representation $\rho\colon \mathrm{GL}(N, \mathbb{C}) \to \mathrm{End}(L)$ for which the matrix entries of $\rho(g)$ are complex-analytic functions of the matrix entries of g. Luckily for $\mathrm{GL}(N, \mathbb{C})$ these representations are all algebraic! The constraint $\rho(gh) = \rho(g)\rho(h)$ is so powerful that any complex-analytic solution is actually algebraic. So, the whole story we told for algebraic representations of $\mathrm{GL}(N, \mathbb{C})$ also applies to complex-analytic ones.

$\mathrm{SL}(N, \mathbb{C})$

We can also get representations of the special linear group $\mathrm{SL}(N, \mathbb{C})$ from Young diagrams. Any Young diagram with at most N rows gives an algebraic irrep of $\mathrm{End}(\mathbb{C}^N)$, and composing this with the inclusion

$$\mathrm{SL}(N, \mathbb{C}) \to \mathrm{End}(\mathbb{C}^N)$$

we get an algebraic irrep of $\mathrm{SL}(N, \mathbb{C})$. We get all the algebraic irreps of $\mathrm{SL}(N, \mathbb{C})$ this way. Even better, the irritating fly in the ointment for $\mathrm{GL}(N, \mathbb{C})$, the determinant representation, become trivial for $\mathrm{SL}(N, \mathbb{C})$. So does the inverse of the determinant representation. So, we get a one-to-one correspondence between these two things:

- algebraic irreps of $\mathrm{SL}(N, \mathbb{C})$, up to isomorphism
- Young diagrams with $< N$ rows.

Furthermore, every algebraic representation of $\mathrm{SL}(N, \mathbb{C})$ is a direct sum of algebraic irreps. So, algebraic representations of $\mathrm{SL}(N, \mathbb{C})$ are classified by *finite collections* of Young diagrams with $< N$ rows.

Here we are thinking of $\mathrm{SL}(N, \mathbb{C})$ as a linear algebraic group. We can also think of it as a complex Lie group. However, all its complex-analytic representations are algebraic. So the same classification applies here too.

$\mathrm{U}(N)$

The unitary group $\mathrm{U}(N)$ is different from the classical groups so far, because the equations defining unitarity involve complex conjugation:

$$gg^* = 1$$

so it's not a linear algebraic group over $\mathbb{C}$. Instead it's a linear algebraic group over $\mathbb{R}$. We shall still study its representations on complex vector spaces, but now the interesting ones are the **real-algebraic representations**: those where the matrix entries of $\rho(g)$ are rational functions of the real and imaginary parts of the matrix entries of g.

To get representations of $\mathrm{U}(N)$ it's convenient to use our knowledge of representations of $\mathrm{GL}(N, \mathbb{C})$. We can take any algebraic irrep of $\mathrm{GL}(N, \mathbb{C})$ and compose it with the inclusion

$$\mathrm{U}(N) \to \mathrm{GL}(N, \mathbb{C})$$

to get a real-algebraic representation of $\mathrm{U}(N)$. The result is an irrep, and we get all the real-algebraic irreps of $\mathrm{U}(N)$ on complex vector spaces this way. In fact, the classification of these real-algebraic irreps of $\mathrm{U}(N)$ completely matches the classification of algebraic irreps of $\mathrm{GL}(N, \mathbb{C})$. We thus get a one-to-one correspondence between these things:

- real-algebraic irreps of $\mathrm{U}(N)$ on complex vector spaces, up to isomorphism
- pairs consisting of a Young diagram with $< N$ rows and an integer.

Furthermore, every real-algebraic representation of $\mathrm{U}(N)$ is a direct sum of real-algebraic irreps.

Alternatively, we can think of $\mathrm{U}(N)$ as a **Lie group**: a group in the category of manifolds (ordinary real manifolds, not complex manifolds). For a Lie group it's natural to study **smooth representations**: those where the matrix entries of $\rho(g)$ are smooth functions of the matrix entries of g. Or we can go further and think of $\mathrm{U}(N)$ as a mere **topological group**: a group in the category of topological spaces. For a topological group it's natural to study **continuous representations**, where the matrix entries of $\rho(g)$ are continuous functions of the matrix entries of g.

But something very nice is true: every smooth representation of $\mathrm{U}(N)$ is automatically real-algebraic, and every continuous representation of *any* Lie group is automatically smooth! So we do not gain any generality by considering smooth or continuous irreps of $\mathrm{U}(N)$: they are both classified by pairs consisting of a Young diagram with $< N$ rows and an integer.

Another variant also turns out to work the same way. In quantum physics we use unitary representations on Hilbert spaces. A finite-dimensional Hilbert space, which is the only kind we'll consider here, is just a finite-dimensional complex vector space with an inner product. A **unitary representation** of a group G on a Hilbert space H is a representation $\rho\colon G \to \mathrm{End}(V)$ such that each of the transformations $\rho(g)$ is unitary.

It turns out that because $\mathrm{U}(N)$ is compact, we can take any continuous representation $\rho\colon \mathrm{U}(N) \to \mathrm{End}(V)$, pick any inner product on the vector space V, and "average it" over the action of $U(N)$ to get a new improved inner product with

$$\langle \rho(g)v, \rho(g)w \rangle = \langle v, w \rangle \quad \text{for all } v, w \in V \text{ and } g \in \mathrm{U}(N).$$

This says that all the transformations $\rho(g)$ are unitary:

$$\rho(g)^* \rho(g) = 1.$$

So, ρ has been promoted to a unitary representation.

Putting this together with what we already have, one can show there is a one-to-one correspondence between these things:

- continuous unitary irreps of $\mathrm{U}(N)$, up to isomorphism
- pairs consisting of a Young diagram with $< N$ rows and an integer.

Also, every continuous unitary representation of $\mathrm{U}(N)$ is a direct sum of continuous unitary irreps.

$\mathbf{SU}(\boldsymbol{N})$

Finally we turn to the special unitary group $\mathrm{SU}(N)$. Since all the main patterns have been laid out, we will go faster now—as usual, not proving things but at least trying to make them plausible. Just as $GL(N, \mathbb{C})$ helps us understand $\mathrm{U}(N)$, $\mathrm{SL}(N, \mathbb{C})$ helps us understand $\mathrm{SU}(N)$. The reason, ultimately, is that $\mathrm{U}(N)$ is the"compact real form" of the complex Lie group $\mathrm{GL}(N, \mathbb{C})$, and $\mathrm{SU}(N)$ is the compact real form of $\mathrm{SL}(N, \mathbb{C})$. But to understand this, one needs to get into Lie theory more deeply than we intend to here.

We can take any algebraic irrep of $\mathrm{SL}(N, \mathbb{C})$ and compose it with the inclusion

$$\mathrm{SU}(N) \to \mathrm{GL}(N, \mathbb{C})$$

to get a representation of $\mathrm{SU}(N)$. This is a real-algebraic irrep, and we get all the real-algebraic irreps of $\mathrm{SU}(N)$ this way. With help from our classification of algebraic irreps of $\mathrm{SL}(N, \mathbb{C})$, we we can show there is a one-to-one correspondence between these things:

- real-algebraic irreps of $\mathrm{SU}(N)$, up to isomorphism
- Young diagrams with $< N$ rows.

Then, by the averaging trick mentioned already for $\mathrm{U}(N)$, we also get a one-to-one correspondence between these things:

- continuous unitary irreps of $\mathrm{SU}(N)$, up to isomorphism
- Young diagrams with $< N$ rows.

Further more, as we have come to expect, in both the real-algebraic case and the continuous unitary case every representation of the given sort is a direct sum of irreps of that sort.

Summary and further directions

Let's summarize what we have seen—but also say a bit more. While we have studied representations on finite-dimensional vector spaces over $\mathbb{C}$, most of the *purely algebraic* results hold for any field of characteristic zero! Fields with nonzero characteristic behave very differently, and in fact the irreducible representations of S_n still haven't been classified over finite fields. But the items with check marks here hold if we replace $\mathbb{C}$ with any field of characteristic zero:

✓ Irreps of S_n are classified by Young diagrams with n boxes.

✓ Polynomial irreps of $\mathrm{End}(\mathbb{C}^N)$ are classified by Young diagrams with $\leq N$ rows.

✓ Polynomial irreps of $\mathrm{GL}(N, \mathbb{C})$ are classified by Young diagrams with $\leq N$ rows.

✓ Algebraic irreps of $\mathrm{GL}(N, \mathbb{C})$ are classified by pairs consisting of a Young diagram with $< N$ rows and an integer.

✓ Algebraic irreps of $\mathrm{SL}(N, \mathbb{C})$ are classified by Young diagrams with $< N$ rows.

- Analytic irreps of $\mathrm{SL}(N, \mathbb{C})$ are classified by Young diagrams with $< N$ rows.

- Analytic irreps of $\mathrm{GL}(N, \mathbb{C})$ are classified by pairs consisting of a Young diagram with $< N$ rows and an integer.
- Real-algebraic irreps of $\mathrm{U}(N)$ are classified by pairs consisting of a Young diagram with $< N$ rows and an integer.
- Continuous unitary irreps of $\mathrm{U}(N)$ are classified by pairs consisting of a Young diagram with $< N$ rows and an integer.
- Real-algebraic irreps of $\mathrm{SU}(N)$ are classified by Young diagram with $< N$ rows.
- Continuous unitary irreps of $\mathrm{SU}(N)$ are classified by Young diagram with $< N$ rows.

But this is far from the end of the story! First of all, we can use n-box Young diagrams packed with numbers $1, \ldots, n$, called **Young tableaux**, to do all sorts of calculations involving irreps of classical groups.

Say we want to figure out the dimension of the irrep of S_n corresponding to some Young diagram. Then we just count the **standard Young tableaux** of that shape: that is, Young tableaux where the numbers increase as we go down any column or across any row. For example, there are two standard Young tableaux of this shape:

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}$$

so this Young diagram:

$$\begin{array}{|c|c|}\hline & \\ \hline & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array}$$

gives a 2-dimensional irrep of S_3.

Or: say we tensor two irreps and want to decompose the result as a direct sum of irreps: how do we do it? We play a little game with Young tableaux and out pops the answer. The relevant buzzword is “Littlewood–Richardson rules”. Or say we have an irrep of S_n and want to know how it decomposes into irreps when we restrict it to a subgroup like S_{n-1}, or similarly for $\mathrm{SL}(N, \mathbb{C})$ and $\mathrm{SL}(N-1, \mathbb{C})$, etc. How do we do this? More messing with Young tableaux. Here one relevant buzzword is “branching rules”.

I’ll warn you right now: there is an *enormous* literature on this stuff. The combinatorics of Young diagrams is one of those things that everyone has worked on, from hardnosed chemists to starry-eyed category theorists. It takes a lifetime to master this material, and I certainly have *not*. But learning even a little is fun, so don’t be *too* scared.

Second of all, Young diagrams are also good for studying the representations of some other classical groups, such as these:

- The **orthogonal group** $\mathrm{O}(N)$, consisting of all orthogonal linear transformations of $\mathbb{R}^N$.
- The **special orthogonal group** $\mathrm{SO}(N)$, consisting of all orthogonal linear transformations of $\mathbb{R}^N$ with determinant 1.
- The **symplectic group** $\mathrm{Sp}(2N)$, consisting of all symplectic linear transformations of $\mathbb{R}^{2N}$.

All these groups have an obvious "tautologous representation", and we can cook up other representations by taking the nth tensor power of this representation and hitting it with minimal idempotents in $\mathbb{C}[S_n]$ coming from Young diagrams. The story I just told you can be repeated with slight or not-so-slight variations for these other groups.

Third, we can "q-deform" the whole story, replacing any one of these classical groups by the associated "quantum group", and replacing $\mathbb{C}[S_n]$ by the corresponding "Hecke algebra". This is really important in topological quantum field theory and the theory of von Neumann algebras.

Fourth, there are nice relationships between Young diagrams and algebraic geometry, like the "Schubert calculus" for the cohomology ring of a Grassmanian.

And there's a lot more, but I have to stop somewhere.

References

I have zipped through a lot of material but not explained it in detail. So, how can you really learn this stuff? If you have a certain amount of patience for old-fashioned terminology, I might recommend going back to the classic text on classical groups:

- Hermann Weyl, *The Classical Groups, Their Invariants and Representations*, Princeton U. Press, Princeton, 1997.

Weyl coined the term "classical groups" for the purposes of this book, which was first published in 1939. His prose is beautiful, but I warn you, this book is not the way to learn Young diagrams in a hurry.

For a user-friendly approach that's aimed at physicists, but still includes proofs of all the key results, you can't beat this:

- Irene Verona Schensted, *A Course on the Applications of Group Theory to Quantum Mechanics*, NEO Press, Box 32, Peaks Island, Maine.

A girlfriend gave me a copy when I was a college student, but only much later did I realize how great a book it is. Unfortunately it's out of print! Someone should reprint this gem. In the meantime, here is another book that covers Young diagrams and their applications to physics:

- Shlomo Sternberg, *Group Theory and Physics*, Cambridge U. Press, Cambridge, 1994.

Both these books, but especially the latter, describe applications of Young diagrams to particle physics, like Gell-Mann's famous "eight-fold way", which was based on positing an $\mathrm{SU}(3)$ symmetry between the up, down and strange quarks.

Then there are more advanced texts, for when your addiction to Young diagrams becomes more serious. For the combinatorial side of things, these are good:

- Gordon Douglas James and Adalbert Kerber, *The Representation Theory of the Symmetric Group*, Addison-Wesley, Reading, Massachusetts, 1981.

- Bruce Eli Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, Springer, Berlin, 2001.

For a more conceptual approach to representation theory that puts Young diagrams in a bigger context, try this:

- Roe Goodman and Nolan R. Wallach, *Representations and Invariants of the Classical Groups*, Cambridge University Press, Cambridge, 1998.

It's sort of an updated version of Weyl's book. And finally, here's a mathematically sophisticated book that really gives you a Young diagram workout:

- William Fulton, *Young Tableaux: With Applications to Representation Theory and Geometry*, Cambridge University Press, Cambridge, 1997.

If you want to learn about Lie groups, these are many good books. I'll list some in rough order of increasing sophistication:

- Brian Hall, *Lie Groups, Lie Algebras, and Representations*, Springer, Berlin, 2003.
- William Fulton and Joe Harris, *Representation Theory — a First Course*, Springer, Berlin, 1991.
- J. Frank Adams, *Lectures on Lie Groups*, University of Chicago Press, Chicago, 2004.
- Sigurdur Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, 1979.
- Daniel Bump, *Lie Groups*, Springer, Berlin, 2004.

The book by Fulton and Harris starts with an introduction to representations of finite groups, especially S_n, and it has a chapter on Young diagrams. For linear algebraic groups, try this:

- James S. Milne, *Basic Theory of Affine Group Schemes*.

Finally, I can't resist mentioning that some friends and I have used category theory to give an updated treatment of some ideas connected to Young diagrams:

- John C. Baez, Joe Moeller and Todd Trimble, Schur functors and categorified plethysm.

The various kinds of representations discussed in the present paper all form categories of a special sort, called "2-rigs". Very roughly speaking, these are categories where we can add using $\oplus$ and multiply using $\otimes$. The free 2-rig on one generator has a remarkable property: irreducible objects in this category are classified by Young diagrams!

This accounts for the prevalence of Young diagrams in the representation theory of classical groups. A classical group G, being the group of linear transformations that preserve some geometrical structure on a vector space V, has V among its representations. There is thus a unique map of 2-rigs from the free 2-rig on one generator to the 2-rig of representations of G sending the generator to V. This map sends Young diagrams to representations of G.

What we've secretly been doing, in these notes, is using this map to classify the irreps of some classical groups.