Basic projective geometry

Teo Banica

Department of Mathematics, University of Cergy-Pontoise, F-95000 Cergy-Pontoise, France. teo.banica@gmail.com

2010 *Mathematics Subject Classification.* 14A10

Key words and phrases. Projective space, Projective manifold

Abstract. This is an elementary introduction to projective geometry. We first discuss basic plane geometry, all good old results going back to the ancient Greeks, and the various simplifications that the projective setting brings, and with a look into higher dimensions too. Then we get into algebraic geometry, of projective flavor, with the Bézout theorem proved, then with the standard algebraic theory developed, and with a look into elliptic curves too. We then discuss the symmetry groups in the projective world, finite or compact, with focus on representation theory, and related diagrams. Finally, we discuss various analytic aspects, regarding projective groups, homogeneous spaces and more general manifolds, notably with various integration techniques.

Preface

Do parallel lines cross? Good question for Humanities, with the answer here varying depending on whom you ask, but with the generally agreed conclusion being that yes, they do cross, say on the grounds that Love will end up uniting them.

So, parallel lines most likely cross, but can we have some understanding of this too, we people in Science. And here, as a first experiment, scientific as they come, wake up in the morning, have a good coffee, I mean just that coffee, and no other things taken by artists and such, and have a look at some railroad tracks. Do that rails cross or not?

And good question this is. Indeed, while you certainly know that the rails won't cross, just imagine the disaster with a train running on that crossing lines, and with this fact being deeply engraved in your brain, well, what you see, with your scientifically trained eyes, rather seems to suggest that the rails will cross. Amazing, isn't it.

In case you doubt, or run into confusion, simply take a picture of that railroad tracks, with this picture being more or less the same thing that your eyes see, independently of what the brain says. And the picture will certainly show that the rails cross.

So, very good, fact established, and the problem is now, can we make now some mathematics, better than the usual mathematics, that we are used to, out of this?

In answer, yes, this is definitely possible, and is something very useful too, for mathematics, and for Science in general. The idea indeed is that the usual mathematics, with usual coordinates and everything, is called "affine", and based on the above observations, with a bit of mathematical effort, we can make it "projective". So, projective geometry and mathematics is what we would like to learn, in order to get smarter.

This book is an elementary introduction to this, projective geometry, and projective mathematics in general. The book is organized in four parts, as follows:

(1) We first discuss basic plane geometry, all good old results going back to the ancient Greeks, and later to the Middle Ages and beyond, and the various simplifications that the projective setting brings, and with a look into higher dimensions too.

(2) Then we get more systematically into algebraic geometry, of projective flavor, that is, study of the projective manifolds, with the Bézout theorem proved, then with the standard algebraic theory developed, and with a look into elliptic curves too.

(3) We discuss then the various symmetry groups in the projective world, that we will usually take to be finite, or even compact, with all sorts of basic theory and examples, and then by focusing on representation theory, and related diagrams.

(4) Finally, we discuss various analytic aspects, regarding projective groups, projective homogeneous spaces, and more general projective manifolds, with some standard differential geometry questions answered, and with various integration techniques.

Many thanks to my math school professors, from the communist Romania of the 1980s, and later to the geometry teachers that I had as a freshman at Bucharest, and later to further geometry teachers, and some teaching colleagues too in Paris, there used to be a fair amount of projective geometry, in our math teaching, at all levels, at that time, and good, pleasant and useful learning that was. Hope one day such things will be back.

Thanks as well to my cats, speaking scientists believing what their eyes see, and drawing quick conclusions out of this, there is no one better than them.

Cergy, March 2025
Teo Banica

Contents

Part I

Projective space

They're really rockin' in Boston
In Pittsburgh, PA
Deep in the heart of Texas
And round the Frisco Bay

CHAPTER 1

Geometry

1a. Parallel lines

Welcome to plane geometry. At the beginner level, which is ours for the moment, this is a story of points and lines. Here is a basic observation, to start with, and we will call this "axiom" instead of "theorem", as the statements which are true and useful are usually called, in mathematics, for reasons that will become clear in a moment:

Axiom 1.1. *Any two distinct points $P \neq Q$ determine a line, denoted PQ.*

Obviously, our axiom holds, and looks like something very useful. Need to draw anything, for various engineering purposes, at your job, or in your garage? The rule will be your main weapon, used exactly as in Axiom 1.1, that is, put the rule on the points $P \neq Q$ that your line must unite, and then draw that line PQ. Actually, in relation with this, we are rather used in practice to draw segments PQ. But in theory, meaning some sort of idealized practice, will having that segment extended to infinity hurt? Certainly not, so this is why our lines PQ in mathematics will be infinite, as above.

Getting now to point, as already announced, why is Axiom 1.1 an axiom, instead of being a theorem? You would probably argue here that this theorem can be proved by using a rule, as indicated above. However, and with my apologies for this, although rock-solid as a scientific proof, this rule thing does not stand as a mathematical proof. This is how things are, you will have to trust me here. And for further making my case, let me mention that my theoretical physics friends agree with me, on the grounds that, when looking with a good microscope at your rule, that rule is certainly bent.

Excuse me, but cat is here, meowing something. So, what is is, cat?

Cat 1.2. *In fact, spacetime itself is bent.*

Okay, thanks cat, so looks like we have multiple problems with the "rule proof" of Axiom 1.1, so that definitely does not qualify as a proof. And so Axiom 1.1 will be indeed an axiom, that is, a true and useful mathematical statement, coming without proof.

Getting now to more discussion, around Axiom 1.1, an interesting question appears in connection with our assumption there $P \neq Q$. Indeed, given a point Q in the plane, we can come up with a sequence of points $P_n \to Q$ vertically, and in this case the lines P_nQ

will all coincide with the vertical at Q. But we can then formally say that the $n \to \infty$ limit of these lines, which makes sense to be denoted QQ, is also the vertical at Q.

However, is this a good idea, or not. The point indeed is that, when doing exactly the same trick with a series of points $P_n \to Q$ horizontally, we will obtain in this way, as our limiting line QQ, the horizontal at Q. Which does not sound very good, but since we seem however to have some sort of valuable idea here, let us formulate:

Job 1.3. *Develop later some kind of analysis theory, generalizing plane geometry, where lines of type QQ make sense too, say as some sort of tangents.*

As a further comment now, still on Axiom 1.1, it is of course understood there that the points $P \neq Q$ appearing there, and the line PQ uniting them, lie in the given plane that we are interested in, in this Part I of the present book. However, Axiom 1.1 obviously holds too in space, and most likely, in higher dimensional spaces too.

So, the question which appears now is, on which type of spaces does Axiom 1.1 hold? And this is a quite interesting question, because if we take a sphere for instance, any two points $P \neq Q$ can be certainly united by a segment, which is by definition the shortest segment, on the sphere, uniting them. And, if we prolong this segment, in the obvious way, what we get is a circle uniting P, Q, that we can call line, and denote P, Q.

However, not so quick. There is in fact a bug with this, because if we take P to be the North Pole, and Q to be the South Pole, any meridian on the globe will do, as PQ. So, as a conclusion, Axiom 1.1 does not really hold on a sphere, but not by much.

Anyway, as before, we seem to have an idea here, so let us formulate:

Job 1.4. *Develop later some kind of advanced geometry theory, generalizing plane geometry, where certain lines PQ can take multiple values.*

And with this, done I guess with the discussion regarding Axiom 1.1, I can only presume that you got as tired of reading this, as I got tired of writing it. Well, this is how things are, geometry is no easy business, and there are certainly plenty of things to be done, and what we will be doing here, based on Axiom 1.1, will be just a beginning.

Excuse me, but cat is meowing again. So, what is it cat, and for God's sake, in the hope that this is not in connection with Axiom 1.1. Please have mercy.

Cat 1.5. *What about a formula of type*

$$PQ = \lambda P + (1 - \lambda)Q$$

proving your axiom.

Okay, thanks cat, but I was already having this in mind, for later in this chapter. So, Axiom 1.1 remains an axiom, please everyone disagreeing with this get out of my math class, and enjoy the sunshine outside. And well, we will see, later in this chapter, how cats and physicists can prove Axiom 1.1, or at least, what their claims are.

Moving ahead now, here is an interesting observation about lines and points in the plane, coming somehow as a complement to Axiom 1.1:

Observation 1.6. *Any two distinct lines $K \neq L$ determine a point, $P = K \cap L$, unless these two lines are parallel, $K||L$.*

So, what do we have here, axiom, theorem, or something else? Not very clear, but on the bottom line, this is something which is certainly true, useful, and provable as before, with a rule. Just carefully draw K, L, and you will certainly get upon $P = K \cap L$.

However, in contrast to Axiom 1.1, there is a bit of a bug with our statement, because we do not know yet, mathematically, what parallel lines means. So, let us formulate:

Definition 1.7. *We say that two lines are parallel, $K||L$, when they do not cross,*

$$K \cap L = \emptyset$$

or when they coincide, $K = L$. Otherwise, we say that K, L cross, and write $K \not\parallel L$.

Here we have tricked a bit, by agreeing to call parallel the pairs of identical lines too, and this for simplifying most of our mathematics, in what follows, trust me here.

As a first remark, with this definition in hand, Observation 1.6 makes now sense, as a formal mathematical statement, and skipping some discussion here, or rather leaving it as an exercise, for reasons which are somewhat clear, we will call this axiom:

Axiom 1.8. *Any two crossing lines $K \not\parallel L$ determine a point, $P = K \cap L$.*

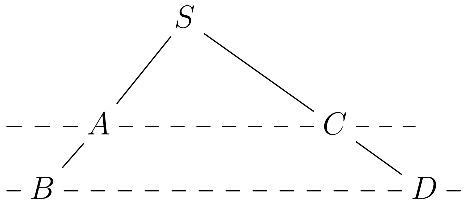
Very good, and now with Axiom 1.1 and Axiom 1.8 in hand, we are potentially ready for doing some geometry. However, this is not exactly true, and we will need as well:

Axiom 1.9. *Given a point not lying on a line, $P \notin L$, we can draw through P a unique parallel to L. That is, we can find a line K satisfying $P \in K$, $K||L$.*

As before, we will leave as an exercise further meditating on all this.

Ready for some math? Here we go, and many things can be said here, especially about parallel lines, which are the main objects of basic geometry. We first have:

Theorem 1.10 (Thales). *Proportions are kept, along parallel lines. That is, given a configuration as follows, consisting of two parallel lines, and of two extra lines,*



the following equality holds:

$$\frac{SA}{SB}=\frac{SC}{SD}$$

Moreover, the converse holds too, in the sense that this implies $AC||BD$.

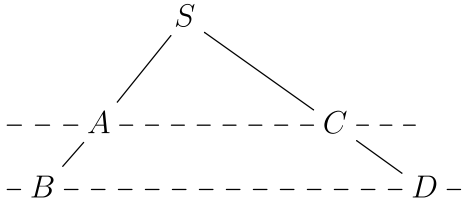
Proof. We have indeed the following computation, based on the usual area formula for the triangles, that is, half of side times height, used multiple times:

$$\begin{aligned}\frac{SA}{SB}&=\frac{area(CSA)}{area(CSB)}\\&=\frac{area(CSA)}{area(CSA)+area(CAB)}\\&=\frac{area(CSA)}{area(CSA)+area(CAD)}\\&=\frac{area(ASC)}{area(ASD)}\\&=\frac{SC}{SD}\end{aligned}$$

As for the converse, we will leave the proof here as an instructive exercise. □

There are some other useful versions of the Thales theorem. First, we have:

Theorem 1.11 (Thales 2). *In the context of the Thales theorem configuration,*

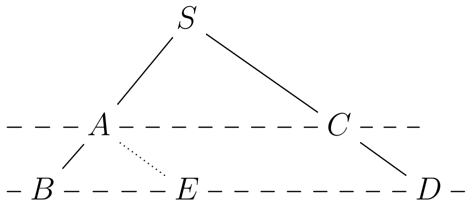


the following equality, involving the same number, holds as well:

$$\frac{SA}{SB}=\frac{AC}{BD}$$

However, the converse of this does not necessarily hold.

Proof. In order to prove the formula in the statement, instead of getting lost into some new area computations, let us draw a tricky parallel, as follows:



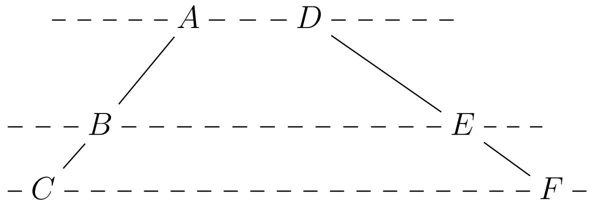
By using Theorem 1.10, we have then the following computation, as desired:

$$\frac{SA}{SB}=\frac{DE}{DB}=\frac{AC}{DB}$$

As for the converse, we will leave the proof here as an instructive exercise. □

As a third Thales theorem now, which is something beautiful too, we have:

Theorem 1.12 (Thales 3). *Given a configuration as follows, consisting of three parallel lines, and of two extra lines, which can cross or not,*



the following equality holds:

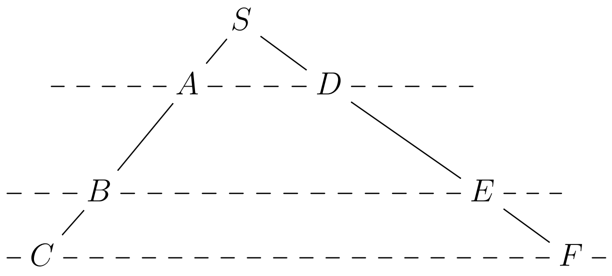
$$\frac{AB}{BC}=\frac{DE}{EF}$$

That is, once again, the proportions are kept, along parallel lines.

Proof. We have two cases here, as follows:

(1) When the two extra lines are parallel, the result is clear, because we have plenty of parallelograms there, and the fractions in question are plainly equal.

(2) When the two lines cross, let us call S their intersection:



Now by using Theorem 1.10 several times, we obtain:

$$\begin{aligned}\frac{AB}{BC}&=\frac{SB-SA}{SC-SB}\\&=\frac{1-\frac{SA}{SB}}{\frac{SC}{SB}-1}\\&=\frac{1-\frac{SD}{SE}}{\frac{SF}{SE}-1}\\&=\frac{SE-SD}{SF-SE}\\&=\frac{DE}{EF}\end{aligned}$$

Thus, we are led to the formula in the statement. □

Importantly, many things can be done with the parallel lines, with a suitably drawn such line hopefully solving, by some kind of miracle, your plane geometry problem.

We will see more illustrations for this general principle in the next section.

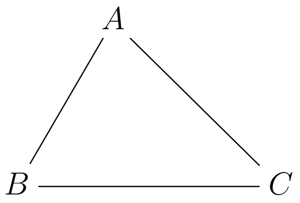
1b. Angles, triangles

Welcome to advanced plane geometry. It all started with triangles, drawn on sand. In order to get started, with some basics, we first have the following key result:

Theorem 1.13. *Given a triangle ABC, the following happen:*

(1) *The angle bisectors cross, at a point called incenter.*
(2) *The medians cross, at a point called barycenter.*
(3) *The perpendicular bisectors cross, at a point called circumcenter.*
(4) *The altitudes cross, at a point called orthocenter.*

Proof. Let us first draw our triangle, with this being always the first thing to be done in geometry, draw a picture, and then thinking and computations afterwards:



Allowing us the freedom to play with some tricks, as advanced mathematicians, both students and professors, are allowed to, here is how the proof goes:

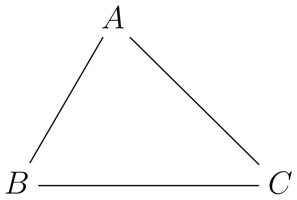
(1) Come with a small circle, inside ABC, and then inflate it, as to touch all 3 edges. The center of the circle will be then at equal distance from all 3 edges, so it will lie on all 3 angle bisectors. Thus, we have constructed the incenter, as required.

(2) This requires different techniques. Let us call $A, B, C \in \mathbb{C}$ the coordinates of A, B, C, and consider the average $P = (A + B + C)/3$. We have then:

$$P = \frac{1}{3} \cdot A + \frac{2}{3} \cdot \frac{B+C}{2}$$

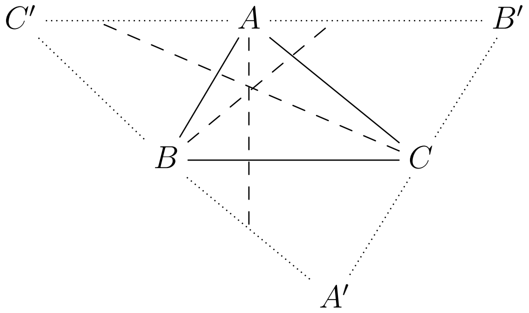
Thus P lies on the median emanating from A, and a similar argument shows that P lies as well on the medians emanating from B, C. Thus, we have our barycenter.

(3) Time to draw a new triangle, for clarity, since we are now on a new page:



Regarding our problem, we can use the same method as for (1). Indeed, come with a big circle, containing ABC, and then deflate it, as for it to pass through A, B, C. The center of the circle will be then at equal distance from all 3 vertices, so it will lie on all 3 perpendicular bisectors. Thus, we have constructed the circumcenter, as required.

(4) This is tougher, and I must admit that, when writing this book, I first struggled a bit with this, then ended looking it up on the internet. So, here is the trick. Draw a parallel to BC at A, and similarly, parallels to AB and AC at C and B. You will get in this way a bigger triangle, upside-down, $A'B'C'$. But then, the circumcenter of $A'B'C'$, that we know to exist from (3), will be the orthocenter of ABC:

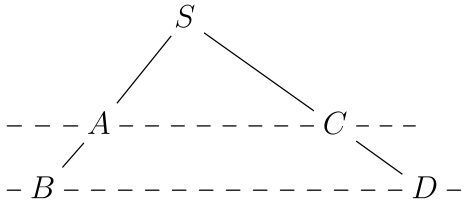


Thus, we are led to the conclusions in the statement. □

Many other things can be said about triangles, and we will be back to this. Importantly, we can now talk about angles, in the obvious way, by using triangles:

Fact 1.14. *We can talk about the angle between two crossing lines, and have some basic theory for the angles going, by using triangles, and Thales, in the obvious way.*

To be more precise here, let us go back to the configuration from the Thales theorem, which was as follows, with two parallel lines, and two other lines:



In this situation, we can say that the two triangles SAC and SBD are similar, and witn an equivalent formulation of similarity being the fact that the angles are equal:

Definition 1.15. *We say that two triangles are similar, and we write*

$$SAC \sim SBD$$

when their respective angles are equal.

The point now is that, in this situation, we can have some mathematics going, for the lengths, coming from the following formula, which is the Thales theorem:

$$\frac{SA}{SB}=\frac{SC}{SD}=\frac{AC}{BD}$$

At the philosophical level now, you might wonder of course what the values of these angles, that we have been heavily using in the above, should be, say as real numbers. But this is something quite tricky, that will take us some time to understand. In the lack of something bright, for the moment, let us formulate the following definition:

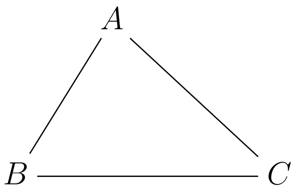
Definition 1.16. *We can talk about the numeric value of angles, as follows:*

(1) *The right angle has value* $90°$*.*
(2) *We can double angles, in the obvious way.*
(3) *Thus, the half right angle has value* $45°$*, and the flat angle has value* $180°$*.*
(4) *We can also triple, quadruple and so on, again in the obvious way.*
(5) *Thus, we can talk about arbitrary rational multiples of* $90°$*.*
(6) *And, with a bit of analysis helping, we can in fact measure any angle.*

So, this will be our starting definition for the numeric values of the angles. Of course, all this might seem a bit improvized, but do not worry, we will come back later to this, with a better, more advanced definition for these numeric values of the angles.

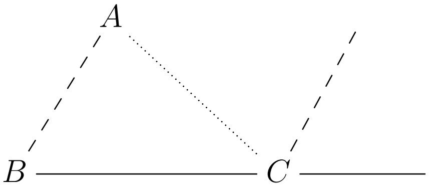
Getting back to work now, theorems and proofs, in relation with the above, here is a key result, which will be our main tool for the study of the angles:

Theorem 1.17. *In an arbitrary triangle*



the sum of all three angles is $180°$*.*

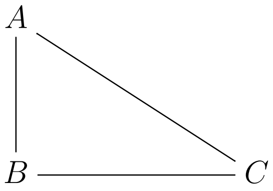
Proof. This does not seem obvious to prove, with bare hands, but as usual, in such situations, some tricky parallels can come to the rescue. Let us prolong indeed the segment BC a bit, on the C side, and then draw a parallel at C, to the line AB, as follows:



But now, we can see that the three angles around C, summing up to the flat angle $180°$, are in fact the 3 angles of our triangle. Thus, theorem proved, just like that. □

Going ahead now with our study of angles, as a continuation of the above, let us first talk about the simplest angle of them all, which is the right angle, denoted $90°$. In relation with it, let us formulate the following definition, making the link with triangles:

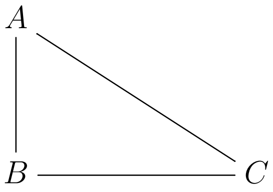
Definition 1.18. *We call right triangle a triangle of type*



having one of the angles equal to $90°$*.*

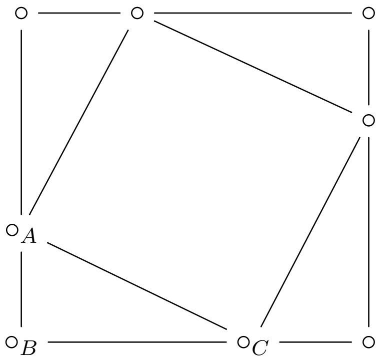
Many things can be said about right triangles, in particular with:

Theorem 1.19 (Pythagoras). *In a right triangle* ABC*,*



we have $AB^2 + BC^2 = AC^2$*.*

PROOF. This comes from the following picture, consisting of two squares, and four triangles which are identical to ABC, as indicated:



Indeed, let us compute the area S of the outer square. This can be done in two ways. First, since the side of this square is $AB+BC$, we obtain:

$$\begin{aligned} S &= (AB+BC)^2 \\ &= AB^2+BC^2+2\times AB\times BC \end{aligned}$$

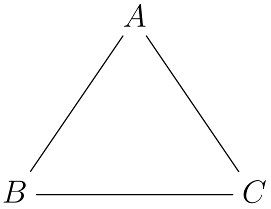
On the other hand, the outer square is made of the smaller square, having side AC, and of four identical right triangles, having sizes AB, BC. Thus:

$$\begin{aligned} S &= AC^2+4\times\frac{AB\times BC}{2} \\ &= AC^2+2\times AB\times BC \end{aligned}$$

Thus, we are led to the conclusion in the statement. □

As a second important angle, we have the 60° angle, which usually appears via:

THEOREM 1.20. *In an equilateral triangle, having all sides equal,*

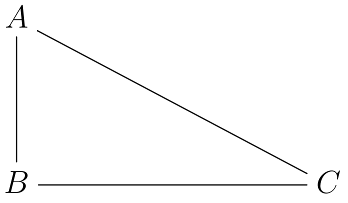


all angles equal 60°*.*

PROOF. This is clear indeed from the fact that the sum is 180°. □

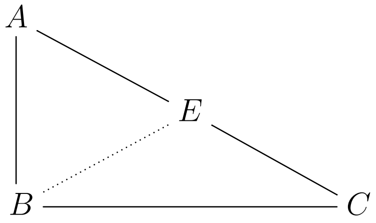
Another interesting angle is the 30° one. About it, we have:

THEOREM 1.21. *In a right triangle having small angles* $30°, 60°$,



we have $AB = AC/2$.

PROOF. This is clear by drawing an equilateral triangle, as follows:



Thus, we are led to the conclusion in the statement. □

We will be back to such things later, when doing trigonometry.

1c. Advanced results

Moving ahead now, many other things can be said about points and lines, and sometimes parallel lines, as a continuation of the Thales theorem. We first have:

THEOREM 1.22 (Desargues). *Two triangles are in perspective axially if and only if they are in perspective centrally.*

PROOF. This is indeed clear in 3D, and the 2D case follows from this. Importantly, as in many other of the results above, there are many cases here, depending on whether various lines cross or not. We will see later how projective geometry simplifies this. □

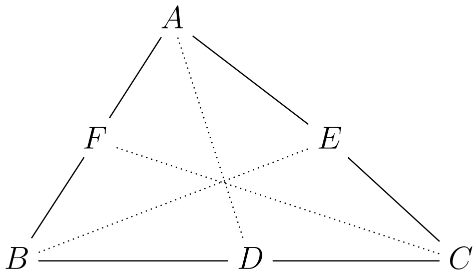
We have as well the following result, going back in time, to Pappus:

THEOREM 1.23 (Pappus). *Given a hexagon with both the odd and the even vertices being colinear, the pairs of opposite sides cross into three colinear points.*

PROOF. This is related to Desargues, and can be proved via several methods. As before with Desargues and other results, there are many cases, depending on whether various lines cross or not. We will see later how projective geometry simplifies this. □

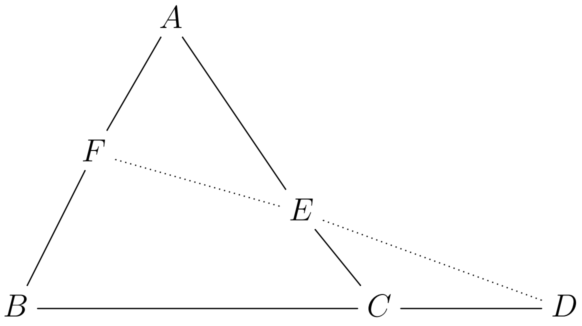
Many other things can be said, about points and lines. We will be back to this.

Let us go back now to basic triangle geometry and centers, as developed before in this chapter. In order to further build on that material, and systematically look at triangle centers, we would like to have general crossing results, of the following type:



We will discusss this slowly, with several results on this subject, and on related topics. First on our list we have the following key result, due to Menelaus:

THEOREM 1.24 (Menelaus). *In a configuration of the following type, with a triangle ABC cut by a line FED,*



we have the following formula, with all segments being taken oriented:

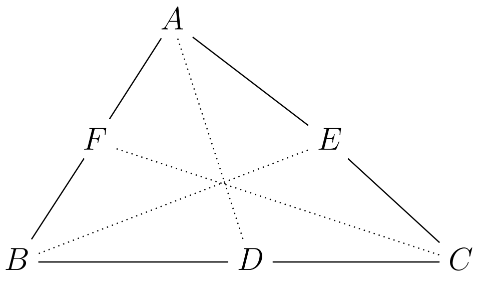
$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=-1$$

Moreover, the converse holds, with this formula guaranteeing that F,E,D are colinear.

PROOF. This is indeed something very standard, by drawing some altitudes. As for the converse, this follows from the main result, in the obvious way. □

We can now answer our original question about crossing lines inside a triangle, drawn from the vertices, with the following remarkable result, due to Ceva:

Theorem 1.25 (Ceva). *In a configuration of the following type, with a triangle ABC containing inner lines AD, BE, CF which cross,*



we have the following formula:

$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=1$$

Moreover, the converse holds, with this formula guaranteeing that AD, BE, CF cross.

Proof. This is indeed something very standard again, which is obviously related to the previous theorem of Menelaus, and which is best seen by computing some areas. As for the converse, this follows from the main result, in the obvious way. □

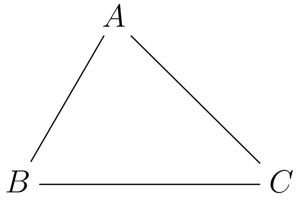
As a basic application of the Ceva theorem, we have now a new point of view on the barycenter. Indeed, the fact that the medians of a triangle cross can be seen as coming from the Ceva theorem, via the following trivial computation:

$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=1\times 1\times 1=1$$

Which is very nice, but needless to say, there is still a lot of work to be done, on the barycenter, in order to understand what cats and physicists know about it, in relation with what was said in the beginning of this chapter. More on this later in this book.

At a more advanced level now, we have the following key result:

Theorem 1.26. *Besides the 4 main centers of a triangle, discussed in the above, many more remarkable points can be associated to a triangle ABC,*



and most of these lie on a line, called Euler line of ABC.

Proof. This is something more technical, which can be proved as well, via some work, the idea with this being as follows:

(1) To start with, it is possible to prove, via some tricks and computations, that the barycenter, the circumcenter and the orthocenter of a triangle are colinear. With this being a key result, among others providing a definition for the Euler line.

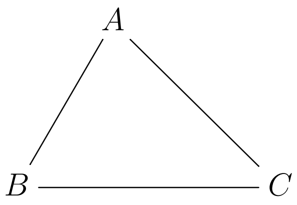
(2) Needless to say, in order for that Euler line to exist, as defined above, the triangle ABC must be assumed to be not equilateral. As for the basic example, for this, for an isosceles triangle, not equilateral, the Euler line is of course the symmetry axis.

(3) At a more advanced level now, as indicated in the statement, it is possible to construct other interesting centers of a triangle, which usually lie on the Euler line. We will be back to this in the next theorem, when discussing the nine-point circle.

(4) Finally, again at the level of more advanced results, we have the question of understanding how these various points lie on the Euler line, meaning understanding the ratios between the distances between them. Again, many things can be said here. □

Along the same lines, we have as well the following result:

Theorem 1.27. *Associated to a triangle ABC,*



we have as well a nine-point circle, whose center lies on the Euler line.

Proof. Again, this is something more technical, which can be proved as well. □

So long for triangles and their centers. This was a very fashionable business long ago, but in more modern times the goals of mathematicians have slightly deviated towards arithmetic, with the must-do thing, instead of constructing a new triangle center, being that of joining the list of generalizators of the Legendre symbol.

As for the truly modern times, here the story is more complicated, with the ultimate goal being that of having your own version of quantum field theory.

For the rest, as already mentioned on several occasions, the above classical geometry material has a number of weaknesses. We will fix this, gradually, in what follows.

1d. Coordinates

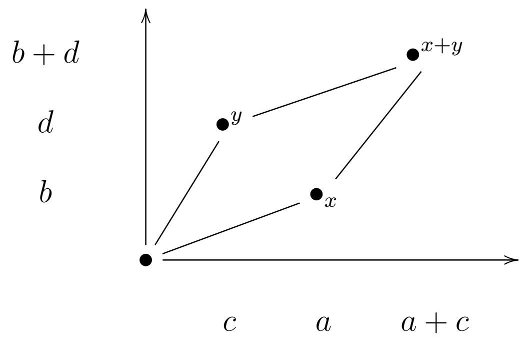
At a more advanced level now, many things from plane geometry can be understood by using coordinates, with each point $x\in\mathbb R^2$ being written as a vector, as follows:

$$x=\begin{pmatrix}a\\ b\end{pmatrix}$$

Of particular interest is the summing operation for such vectors, which, according to the usual calculus rules for the vectors, is given by the following formula:

$$x=\begin{pmatrix}a\\ b\end{pmatrix}\ ,\ y=\begin{pmatrix}c\\ d\end{pmatrix}\implies x+y=\begin{pmatrix}a+c\\ b+d\end{pmatrix}$$

Indeed, as you surely know well from calculus, geometrically, the idea here is simply that the vectors add by forming a parallelogram, as follows:



In practice, the summing operation is usefully complemented by the multiplication by scalars operation, which is given by the following very intuitive formula:

$$x=\begin{pmatrix}a\\ b\end{pmatrix}\implies\lambda x=\begin{pmatrix}\lambda a\\ \lambda b\end{pmatrix}$$

Finally, of particular interest too, in relation with the computation of the lengths, is the following formula, allowing us to compute the length of any vector:

$$x=\begin{pmatrix}a\\ b\end{pmatrix}\implies||x||=\sqrt{a^2+b^2}$$

Very good, and time now to see how our coordinate technology works, if that is worth something, or not. We will review here all the triangle and basic geometry material from before, with new proofs for everything, using coordinates, no less than that.

So, God bless, and let us get started. As a first good surprise, in what regards the axiomatics from the beginning of this chapter, that is literally nuked by coordinates.

We first have, indeed, regarding the first axiom of geometry, that we started this book with, the following theorem, coming along with a trivial proof:

Theorem 1.28. *Any two distinct points $P \neq Q$ determine a line, denoted PQ.*

Proof. This is clear indeed, with coordinates, because we have:

$$PQ = \lambda P + (1-\lambda)Q$$

So, very good news, axiom becoming theorem, what more can we wish for. □

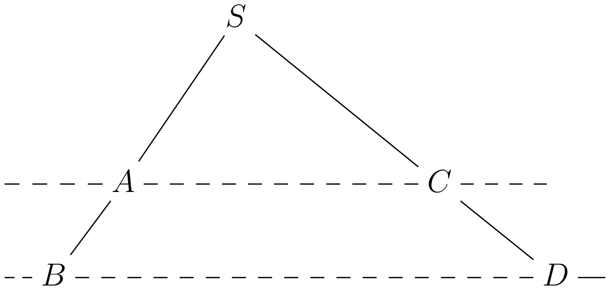
Same situation for the second axiom, which becomes a theorem too:

Theorem 1.29. *Given a point not lying on a line, $P \notin L$, we can draw through P a unique parallel to L. That is, we can find a line K satisfying $P \in K$, $K||L$.*

Proof. This is again clear with coordinates. □

Getting now to the next thing that we did before, namely the Thales theorem, and as further good news, that drastically simplifies with coordinates, as follows:

Theorem 1.30 (Thales). *Proportions are kept, along parallel lines. That is, given a configuration as follows, consisting of two parallel lines, and of two extra lines,*



the following equality holds:

$$\frac{SA}{SB} = \frac{SC}{SD}$$

Moreover, the converse of this holds too, in the sense that, in the context of a picture as above, if this equality is satisfied, then the lines AC and BD must be parallel.

Proof. Again, this is clear with coordinates, and in fact the other formulations of the Thales theorem, also from Part I, are clear as well too, again with coordinates. To be more precise, for the above configuration, the conclusion is as follows:

$$\frac{SA}{SB} = \frac{SC}{SD} = \frac{AC}{BD}$$

In addition, we can prove Thales 3 as well, again using coordinates. □

Next, we have the Desargues theorem:

Theorem 1.31 (Desargues). *Two triangles are in perspective axially if and only if they are in perspective centrally.*

Proof. Again, this is clear with coordinates. □

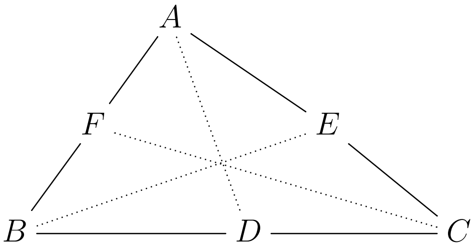
Next, we have the Pappus theorem:

Theorem 1.32 (Pappus). *Given a hexagon with both the odd and the even vertices being colinear, the pairs of opposite sides cross into three colinear points.*

Proof. Again, this is clear with coordinates. □

Getting now to the barycenter theorem, this drastically simplifies, as follows:

Theorem 1.33 (Barycenter). *Given a triangle ABC, its medians cross,*



at a point called barycenter, lying at $1/3-2/3$ on each median.

Proof. Let us call $A,B,C\in\mathbb{R}^2$ the coordinates of the vertices A,B,C, and consider the average $P=(A+B+C)/3$. We have then:

$$P=\frac{1}{3}\cdot A+\frac{2}{3}\cdot\frac{B+C}{2}$$

Thus P lies on the median emanating from A, and a similar argument shows that P lies as well on the medians emanating from B,C. Thus, we have our barycenter. □

We can prove now as well some other things claimed before, as follows:

Theorem 1.34. *The gravity center of a triangle ABC is as follows:*

1. *In the 0-dimensional case, that is, when putting equal weigths at the vertices A,B,C, and computing the center, this is the barycenter.*
2. *In the 1-dimensional case, that is, with the sides AB,BC,AC have weigths proportional with their length, this is, in general, different from the barycenter.*
3. *In the 2-dimensional case, that is, with the triangle ABC itself, as an area, having a weight, uniformly distributed, this is again the barycenter.*

Proof. Again, this is clear with coordinates. Indeed, (1) is something which follows from the proof of Theorem 1.33, then (2) follows from an easy computation, and (3) is something which is elementary too, with a bit of analysis know-how. □

Getting now to the other centers of a triangle, we have here:

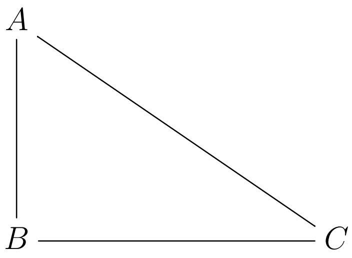
THEOREM 1.35. *Given a triangle ABC, the following happen:*

(1) *The angle bisectors cross, at a point called incenter.*
(2) *The perpendicular bisectors cross, at a point called circumcenter.*
(3) *The altitudes cross, at a point called orthocenter.*

PROOF. Again, such things can be proved with coordinates, and patience. We will actually leave some of the calculations here as an instructive exercise for you, reader. □

Coming next, we have the theorem of Pythagoras:

THEOREM 1.36 (Pythagoras). *In a right triangle ABC,*

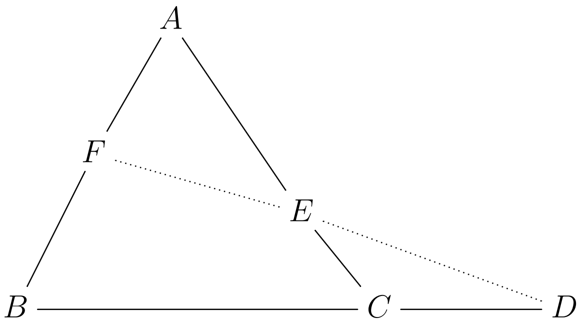


we have $AB^2+BC^2=AC^2$.

PROOF. Again, this is clear with coordinates. □

Next, we have the following key result, due to Menelaus:

THEOREM 1.37 (Menelaus). *In a configuration of the following type, with a triangle ABC cut by a line FED,*



we have the following formula, with all segments being taken oriented:

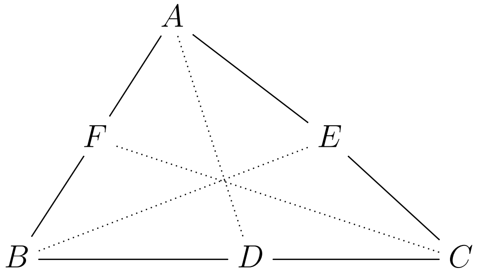
$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=-1$$

Moreover, the converse holds, with this formula guaranteeing that F,E,D are colinear.

PROOF. Again, this is clear with coordinates. □

Next, we have the following remarkable result, due to Ceva:

THEOREM 1.38 (Ceva). *In a configuration of the following type, with a triangle ABC containing inner lines AD, BE, CF which cross,*



we have the following formula:

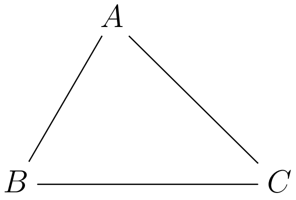
$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=1$$

Moreover, the converse holds, with this formula guaranteeing that AD, BE, CF cross.

PROOF. Again, this is clear with coordinates. □

At a more advanced level now, we have the following key result:

THEOREM 1.39. *Besides the 4 main centers of a triangle, discussed in the above, many more remarkable points can be associated to a triangle ABC,*

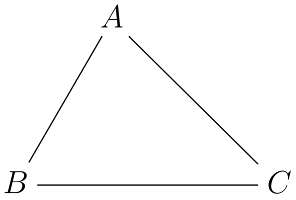


and most of these lie on a line, called Euler line of ABC.

PROOF. Proving this with coordinates is a good exercise for you, reader. □

Along the same lines, we have as well the following result:

THEOREM 1.40. *Associated to a triangle ABC,*



we have as well a nine-point circle, whose center lies on the Euler line.

Proof. Again, proving this with coordinates is a good exercise for you, reader. □

As a conclusion to all this, coordinates seem to perfom quite well, and you might probably have this question right now, why not having started the present book with coordinates. In answer, modesty and patience, this is how math is best learned. We will actually see right next that our present $\mathbb{R}^2$ coordinates can be beaten themselves by some better coordinates, namely the $\mathbb{C}$ ones. So, long story still to go, and ho hurry.

So, let us talk now about complex coordinates. As a starting point, we have:

Theorem 1.41. *The complex numbers, $z = a + ib$ with $a, b \in \mathbb{R}$ and with i being a formal number satisying $i^2 = -1$, form a field $\mathbb{C}$. Moreover:*

1. *We have a field embedding $\mathbb{R} \subset \mathbb{C}$, given by $a \to a + 0 \cdot i$.*
2. *Additively, we have $\mathbb{C} \simeq \mathbb{R}^2$, with $z = a + ib$ corresponding to (a, b).*
3. *The length of vectors $r = |z|$, with $z = a + ib$, is given by $r = \sqrt{a^2 + b^2}$.*
4. *With $z = r(\cos t + i \sin t)$, the products $z = z'z''$ are given by $r = r'r''$, $t = t' + t''$.*
5. *We have $e^{it} = \cos t + i \sin t$, so we can write $z = re^{it}$.*
6. *There are N solutions to the equation $z^N = 1$, called N-th roots of unity.*
7. *Any degree 2 equation with complex coefficients has both roots in $\mathbb{C}$.*

Proof. We have a field, with $z^{-1} = (a - ib)/(a^2 + b^2)$, and regarding the rest:

(1) This is clear.

(2) Again, this is clear.

(3) Again, this is clear. Observe also that we have $r^2 = z\bar{z}$, with $\bar{z} = a - ib$.

(4) We need here the formulae for the sines and cosines of sums, which are as follows, coming from some trigonometry, done the old way, with triangles in the plane:

$$\cos(s + t) = \cos s \cos t - \sin s \sin t$$

$$\sin(s + t) = \sin s \cos t + \cos s \sin t$$

Indeed, with these formulae in hand, we have the following computation, as desired:

$$\begin{aligned}
& (\cos s + i \sin s)(\cos t + i \sin t) \\
= \ & (\cos s \cos t + i^2 \sin s \sin t) + i(\sin s \cos t + \cos s \sin t) \\
= \ & (\cos s \cos t - \sin s \sin t) + i(\sin s \cos t + \cos s \sin t) \\
= \ & \cos(s + t) + i \sin(s + t)
\end{aligned}$$

(5) This follows from some heavy calculus, namely Taylor formula for $\exp, \sin, \cos$:

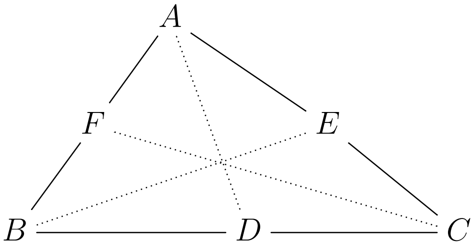
$$\begin{aligned} e^{it} &= \sum_{k=0}^\infty \frac{(it)^k}{k!} \\ &= \sum_{l=0}^\infty \frac{(it)^{2l}}{(2l)!} + \sum_{l=0}^\infty \frac{(it)^{2l+1}}{(2l+1)!} \\ &= \sum_{l=0}^\infty (-1)^l \frac{t^{2l}}{(2l)!} + i \sum_{l=0}^\infty (-1)^l \frac{t^{2l+1}}{(2l+1)!} \\ &= \cos t + i \sin t \end{aligned}$$

(6) This is clear from (5), with $z = w^k$, with $w = e^{2\pi i/N}$ and $k = 0, 1, \ldots, N-1$.

(7) This follows in the usual way, with $\sqrt{re^{it}} = \pm\sqrt{r}e^{it/2}$ at the end, using (5). □

Getting now to geometry, using complex numbers, many things can be said here, and as a sample result, we have a better point of view on the barycenter, as follows:

THEOREM 1.42 (Barycenter). *Given a triangle ABC, its medians cross,*



at a point called barycenter, lying at $1/3 - 2/3$ on each median.

PROOF. Let us call $A, B, C \in \mathbb{C}$ the coordinates of the vertices A, B, C, and consider the average $P = (A + B + C)/3$. We have then:

$$P = \frac{1}{3} \cdot A + \frac{2}{3} \cdot \frac{B + C}{2}$$

Thus P lies on the median emanating from A, and a similar argument shows that P lies as well on the medians emanating from B, C. Thus, we have our barycenter. □

At a more advanced level, many interesting things can be done in relation with orthogonality, which in complex coordinates reads:

$$x \perp y \iff \frac{x}{y} \in i\mathbb{R}$$

Also, the equations of circles are now something simpler, as follows:

$$|x - c| = r$$

As an application of this technology, we have the following result:

Theorem 1.43. *Given a triangle ABC, the following happen:*

(1) *The angle bisectors cross, at a point called incenter.*
(2) *The perpendicular bisectors cross, at a point called circumcenter.*
(3) *The altitudes cross, at a point called orthocenter.*

Proof. Again, such things can be proved with complex coordinates. We will actually leave some of the calculations here as an instructive exercise for you, reader. □

1e. Exercises

Exercises:

Exercise 1.44.

Exercise 1.45.

Exercise 1.46.

Exercise 1.47.

Exercise 1.48.

Exercise 1.49.

Exercise 1.50.

Exercise 1.51.

Bonus exercise.

CHAPTER 2

Projective plane

2a. Projective plane

Welcome to projective geometry. In order to have the parallel lines crossing, which is something that would be desirable, here are some axioms, to start with:

Definition 2.1. *A projective space is a space consisting of points and lines, subject to the following conditions:*

(1) *Each 2 points determine a line.*
(2) *Each 2 lines cross, on a point.*

Obviously, this is something quite general, because a line itself is a projective space, in the above sense. Note that a circle is a projective space too, in the above sense. We will be back to such trivial examples later, when talking further axiomatization.

The main example that we will be interested in, in this chapter, is as follows:

Definition 2.2. *The projective plane, denoted $P^2_\mathbb{R}$, is the space of lines in $\mathbb{R}^3$ passing through the origin. To be more precise:*

(1) *We call each of the lines in $\mathbb{R}^3$ passing through the origin a point of $P^2_\mathbb{R}$.*
(2) *We also call each plane in $\mathbb{R}^3$ passing through the origin a line of $P^2_\mathbb{R}$.*

And in the hope that you will not find this too confusing, but no worries, we will get used to this, which is something quite clever, as we will soon discover.

Getting started with our study, in relation with Definition 2.1, observe that the following happen, in relation with the points and lines of $P^2_\mathbb{R}$, as constructed above:

(1) Each 2 points determine a line. Indeed, 2 points in our sense means 2 lines in $\mathbb{R}^3$ passing through the origin, and these 2 lines obviously determine a plane in $\mathbb{R}^3$ passing through the origin, namely the plane they belong to, which is a line in our sense.

(2) Each 2 lines cross, on a point. Indeed, 2 lines in our sense means 2 planes in $\mathbb{R}^3$ passing through the origin, and these 2 planes obviously determine a line in $\mathbb{R}^3$ passing through the origin, namely their intersection, which is a point in our sense.

As a conclusion to this, what we have is a projective space in the sense of Definition 2.1. Let us record this finding as a theorem, as follows:

Theorem 2.3. *The projective plane $P^2_\mathbb{R}$ is a projective space, in the sense that:*

(1) *Each 2 points determine a line.*
(2) *Each 2 lines cross, on a point.*

Proof. This follows indeed form the above discussion. □

In order to concretely deal now with $P^2_\mathbb{R}$, say by using coordinates, as we would prefer, several methods are available. We first have the following result:

Theorem 2.4. *The projective plane $P^2_\mathbb{R}$ appears, alternatively, as the quotient*

$$P^2_\mathbb{R} = \mathbb{R}^3 - \{0\}/\sim$$

with $\sim$ being the proportionality of vectors, given by $x \sim y$ when $x = \lambda y$, with $\lambda \neq 0$.

Proof. We know that the projective plane $P^2_\mathbb{R}$ appears by definition as the space of lines in $\mathbb{R}^3$ passing through the origin, and this gives the result. □

As a continuation of this, we can restrict if we want the attention to the vectors on the unit sphere $S^2_\mathbb{R} \subset \mathbb{R}^3$, and this because any line in $\mathbb{R}^3$ passing through the origin will certainly cross this sphere. We are led in this way to the following result:

Theorem 2.5. *The projective plane $P^2_\mathbb{R}$ appears also as the quotient*

$$P^2_\mathbb{R} = S^2_\mathbb{R}/\sim$$

with $\sim$ being the proportionality of vectors on the sphere, given by $x \sim y$ when $x = \pm y$.

Proof. According to the discussion above, we can restrict the attention to the vectors on the sphere $S^2_\mathbb{R} \subset \mathbb{R}^3$, and this gives the following formula, with $\sim$ standing as before for the proportionality of vectors in space, given by $x \sim y$ when $x = \lambda y$, with $\lambda \neq 0$:

$$P^2_\mathbb{R} = S^2_\mathbb{R}/\sim$$

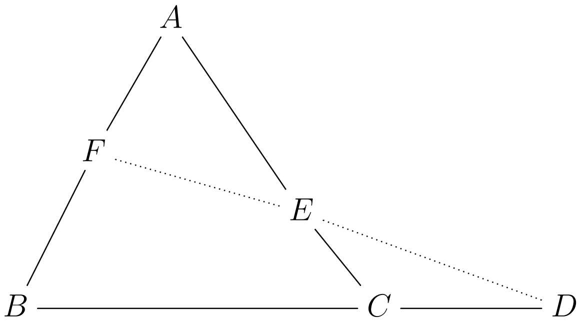
But, it is clear that our line will cross the sphere in exactly two points $\pm x$, and we conclude that we have the formula in the statement. □

Many other things can be said, as a continuation of the above, and notably in relation with the picture for $P^2_\mathbb{R}$ coming from Theorem 2.5. We will be back to this.

2b. Projective geometry

Time now to do some projective geometry. Following the material in chapter 1, let us start with the following key result, due to Menelaus:

Theorem 2.6 (Menelaus). *In a configuration of the following type, with a triangle ABC cut by a line FED,*



we have the following formula, with all segments being taken oriented:

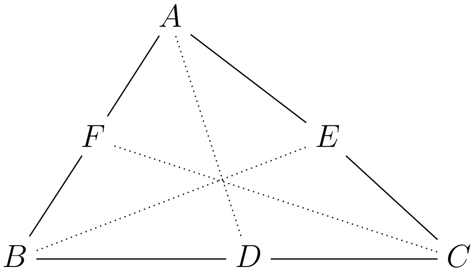
$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=-1$$

Moreover, the converse holds, with this formula guaranteeing that F,E,D are colinear.

Proof. This is indeed best viewed in the projective geometry setting. □

Next, we have the following remarkable result, due to Ceva:

Theorem 2.7 (Ceva). *In a configuration of the following type, with a triangle ABC containing inner lines AD,BE,CF which cross,*



we have the following formula:

$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=1$$

Moreover, the converse holds, with this formula guaranteeing that AD,BE,CF cross.

Proof. Again, this is best viewed in the projective geometry setting. □

Many other things can be said, as a continuation of the above, for instance with some theory for the curves in the projective plane. We will be back to this.

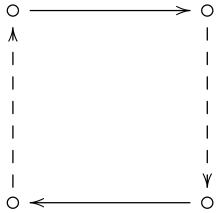
2c. Shape, embeddings

Back now to the projective plane itself, this remains a quite mysterious object. Regarding its shape, we have the following result, formulated of course quite informally:

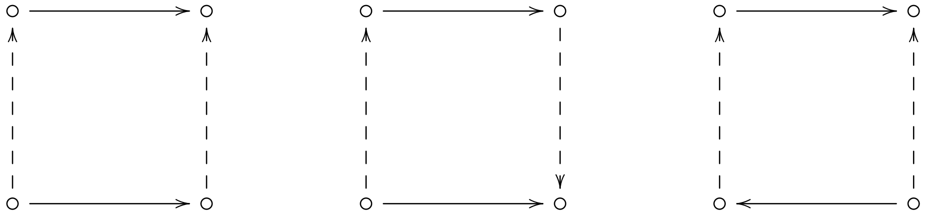
Theorem 2.8. *The projective plane $P^2_\mathbb{R}$ is some sort of twisted sphere.*

Proof. All this is of course a bit informal, the idea being as follows:

(1) We know that $P^2_\mathbb{R}$ corresponds to the upper hemisphere of the sphere $S^2_\mathbb{R}\subset\mathbb{R}^3$, with the points on the equator identified via $x=-x$. Topologically speaking, we can deform if we want the hemisphere into a square, with the equator becoming the boundary of this square, and in this picture, the $x=-x$ identification corresponds to a "identify opposite edges, with opposite orientations" folding method for the square:



(2) Thus, we have our space. In order to understand now what this beast is, let us look first at the other 3 possible methods of folding the square, which are as follows:



Regarding the first space, the one on the left, things here are quite simple. Indeed, when identifying the solid edges we get a cylinder, and then when further identifying the dotted edges, what we get is some sort of closed cylinder, which is a torus.

(3) Regarding the second space, the one in the middle, things here are more tricky. Indeed, when identifying the solid edges we get again a cylinder, but then when further identifying the dotted edges, we obtain some sort of "impossible" closed cylinder, called Klein bottle. This Klein bottle obviously cannot be drawn in 3 dimensions, but with a bit of imagination, you can see it, in its full splendor, in 4 dimensions.

(4) Finally, regarding the third space, the one on the right, we know by symmetry that this must be the Klein bottle too. But we can see this as well via our standard folding method, namely identifying solid edges first, and dotted edges afterwards. Indeed, we first obtain in this way a Möbius strip, and then, well, the Klein bottle.

(5) With these preliminaries made, and getting back now to the projective space $P^2_\mathbb{R}$, we can see that this is something more complicated, of the same type, reminding the torus and the Klein bottle. So, we will call it "sort of twisted sphere", as in the statement, and exercise for you to imagine how this beast looks like, in 4 dimensions. □

All this is quite exciting, and reminds childhood and primary school, but is however a bit tiring for our neurons, guess that is pure mathematics. It is possible to come up with some explicit formulae for the embedding $P^2_\mathbb{R}\subset\mathbb{R}^4$, which are useful in practice, allowing us to do some analysis over $P^2_\mathbb{R}$, and we will leave this as an instructive exercise.

There is some linear algebra to be done here too, by identifying the lines in $\mathbb{R}^3$ with the corresponding rank 1 projections, along with many other things, and we have:

Theorem 2.9. *The projective space $P^2_\mathbb{R}$ can be thought of as being the space of rank 1 projections in the matrix algebra $M_3(\mathbb{R})$, given by*

$$P_x=\frac{1}{||x||^2}(x_ix_j)_{ij}$$

by identifying the lines in $\mathbb{R}^3$ passing through the origin with the corresponding rank 1 projections in $M_3(\mathbb{R})$, in the obvious way.

Proof. There are several things going on here, the idea being as follows:

(1) The main assertion is more or less clear from definitions, the point being that the lines in $\mathbb{R}^3$ passing through the origin are obviously in bijection with the corresponding rank 1 projections. Thus, we obtain the interpretation of $P^2_\mathbb{R}$ in the statement.

(2) Regarding now the formula of the rank 1 projections, which is a must-know, for this, and in everyday life, consider a vector $y\in\mathbb{R}^3$. Its projection on $\mathbb{R}x$ must be a certain multiple of x, and we are led in this way to the following formula:

$$P_xy=\frac{<y,x>}{<x,x>}x=\frac{1}{||x||^2}<y,x>x$$

(3) But with this in hand, we can now compute the entries of P_x, as follows:

$$\begin{aligned}(P_x)_{ij}&=<P_xe_j,e_i>\\&=\frac{1}{||x||^2}<e_j,x><x,e_i>\\&=\frac{x_jx_i}{||x||^2}\end{aligned}$$

Thus, we are led to the formula in the statement. □

Regarding now embeddings of $P^2_\mathbb{R}$ into Euclidean spaces $\mathbb{R}^n$, many things can be said, with a straightforward construction here being as follows:

THEOREM 2.10. *The projective space $P^2_{\mathbb{R}}$ is a smooth manifold, with charts*

$$(x_1,x_2,x_3)\to\left(\frac{x_1}{x_i},\ldots,\frac{x_{i-1}}{x_i},\frac{x_{i+1}}{x_i},\ldots,\frac{x_3}{x_i}\right)$$

where $x_i\neq 0$. This manifold is compact, and of dimension 2*.*

PROOF. We know that $P^2_{\mathbb{R}}$ appears as the space of lines in $\mathbb{R}^3$ passing through the origin, so we have the following formula, with $\sim$ being the proportionality of vectors, given as usual by $x\sim y$ when $x=\lambda y$, for some scalar $\lambda\neq 0$:

$$P^2_{\mathbb{R}}=\mathbb{R}^3-\{0\}/\sim$$

With this discussion made, let us get now to what is to be proved. Obviously, once we fix an index $i\in\{1,\ldots,3\}$, the condition $x_i\neq 0$ on the vectors $x\in\mathbb{R}^3-\{0\}$ defines an open subset $U_i\subset P^2_{\mathbb{R}}$, and the open subsets that we get in this way cover $P^2_{\mathbb{R}}$:

$$P^2_{\mathbb{R}}=U_1\cup\ldots\cup U_3$$

Moreover, the map in the statement is injective $U_i\to\mathbb{R}^2$, and it is clear too that the changes of charts are C^∞. Thus, we have our smooth manifold, as claimed. □

Many other things can be said about $P^2_{\mathbb{R}}$, and we will be back to this.

2d. Curves and more

Curves and more.

2e. Exercises

Exercises:

EXERCISE 2.11.

EXERCISE 2.12.

EXERCISE 2.13.

EXERCISE 2.14.

EXERCISE 2.15.

EXERCISE 2.16.

EXERCISE 2.17.

EXERCISE 2.18.

Bonus exercise.

CHAPTER 3

Projective space

3a. Projective space

Welcome to projective geometry, this time in 3 or more dimensions. In order to have the parallel lines crossing, we use the same axioms as before, namely:

DEFINITION 3.1. *A projective space is a space consisting of points and lines, subject to the following conditions:*

(1) *Each* 2 *points determine a line.*
(2) *Each* 2 *lines cross, on a point.*

As noted before in chapter 2, this formalism is something quite general, because a line itself is a projective space, in the above sense. Note that a circle is a projective space too, in the above sense. More on these trivial examples in a moment.

The main example that we will be interested in, in this chapter, is as follows:

DEFINITION 3.2. *The real projective space, denoted $P^{N-1}_\mathbb{R}$, is the space of lines in $\mathbb{R}^N$ passing through the origin. To be more precise:*

(1) *We call each of the lines in $\mathbb{R}^N$ passing through the origin a point of $P^{N-1}_\mathbb{R}$.*
(2) *We also call each plane in $\mathbb{R}^N$ passing through the origin a line of $P^{N-1}_\mathbb{R}$.*

Getting started with this, in relation with Definition 3.1, observe that the following happen, in relation with the points and lines of $P^{N-1}_\mathbb{R}$, as constructed above:

(1) Each 2 points determine a line. Indeed, 2 points in our sense means 2 lines in $\mathbb{R}^N$ passing through the origin, and these 2 lines obviously determine a plane in $\mathbb{R}^N$ passing through the origin, namely the plane they belong to, which is a line in our sense.

(2) Each 2 lines cross, on a point. Indeed, 2 lines in our sense means 2 planes in $\mathbb{R}^N$ passing through the origin, and these 2 planes obviously determine a line in $\mathbb{R}^N$ passing through the origin, namely their intersection, which is a point in our sense.

Let us record this finding as a theorem, as follows:

THEOREM 3.3. *The space $P^{N-1}_\mathbb{R}$ is indeed a projective space, in the sense that each* 2 *points determine a line, and each* 2 *lines cross, on a point.*

Proof. This follows indeed form the above discussion. □

In order to concretely deal now with $P^{N-1}_\mathbb{R}$, say by using coordinates, as we would prefer, several methods are available. We first have the following result:

Theorem 3.4. *The projective plane $P^{N-1}_\mathbb{R}$ appears, alternatively, as the quotient*

$$P^{N-1}_\mathbb{R}=\mathbb{R}^N-\{0\}/\sim$$

with $\sim$ being the proportionality of vectors, given by $x\sim y$ when $x=\lambda y$, with $\lambda\neq 0$.

Proof. We know that the projective plane $P^{N-1}_\mathbb{R}$ appears by definition as the space of lines in $\mathbb{R}^N$ passing through the origin, and this gives the result. □

As a continuation of this, we can restrict if we want the attention to the vectors on the unit sphere $S^{N-1}_\mathbb{R}\subset\mathbb{R}^N$, and this because any line in $\mathbb{R}^N$ passing through the origin will certainly cross this sphere. We are led in this way to the following result:

Theorem 3.5. *The projective plane $P^{N-1}_\mathbb{R}$ appears also as the quotient*

$$P^{N-1}_\mathbb{R}=S^{N-1}_\mathbb{R}/\sim$$

with $\sim$ being the proportionality of vectors on the sphere, given by $x\sim y$ when $x=\pm y$.

Proof. According to the discussion above, we can restrict the attention to the vectors on the sphere $S^{N-1}_\mathbb{R}\subset\mathbb{R}^N$, and this gives the following formula, with $\sim$ standing as before for the proportionality of vectors in space, given by $x\sim y$ when $x=\lambda y$, with $\lambda\neq 0$:

$$P^2_\mathbb{R}=S^2_\mathbb{R}/\sim$$

But, it is clear that our line will cross the sphere in exactly two points $\pm x$, and we conclude that we have the formula in the statement. □

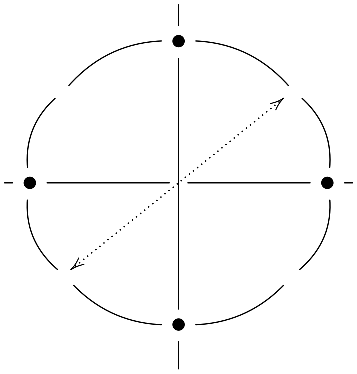
Many other things can be said, as a continuation of the above, and notably in relation with the picture for $P^{N-1}_\mathbb{R}$ coming from Theorem 3.5. To be more precise, we have:

Theorem 3.6. *In small dimensions, the projective space $P^{N-1}_\mathbb{R}$ is as follows:*

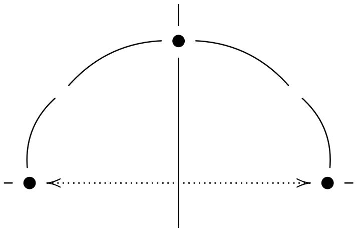
(1) *$P^1_\mathbb{R}$ is the usual circle.*
(2) *$P^2_\mathbb{R}$ is some sort of twisted sphere.*

Proof. We have several assertions here, with all this being of course a bit informal, and self-explanatory, the idea and some further details being as follows:

(1) At $N=2$, a line in $\mathbb{R}^2$ passing through the origin corresponds to 2 opposite points on the unit circle $\mathbb{T}\subset\mathbb{R}^2$, according to the following scheme:



Thus, $P^1_{\mathbb{R}}$ corresponds to the upper semicircle of $\mathbb{T}$, with the endpoints identified, and so we obtain a circle, $P^1_{\mathbb{R}}=\mathbb{T}$, according to the following scheme:



(2) At $N=3$, this is something that we already know, from chapter 2. □

There is some linear algebra to be done here too, by identifying the lines in $\mathbb{R}^N$ with the corresponding rank 1 projections, along with many other things, and we have:

Theorem 3.7. *The projective space $P^{N-1}_{\mathbb{R}}$ can be thought of as being the space of rank 1 projections in the matrix algebra $M_N(\mathbb{R})$, given by*

$$P_x=\frac{1}{||x||^2}(x_ix_j)_{ij}$$

by identifying the lines in $\mathbb{R}^N$ passing through the origin with the corresponding rank 1 projections in $M_N(\mathbb{R})$, in the obvious way.

Proof. There are several things going on here, the idea being as follows:

(1) The main assertion is more or less clear from definitions, the point being that the lines in $\mathbb{R}^N$ passing through the origin are obviously in bijection with the corresponding rank 1 projections. Thus, we obtain the interpretation of $P^{N-1}_{\mathbb{R}}$ in the statement.

(2) Regarding now the formula of the rank 1 projections, which is a must-know, for this, and in everyday life, consider a vector $y\in\mathbb R^N$. Its projection on $\mathbb Rx$ must be a certain multiple of x, and we are led in this way to the following formula:

$$P_xy=\frac{<y,x>}{<x,x>}x=\frac{1}{||x||^2}<y,x>x$$

(3) But with this in hand, we can now compute the entries of P_x, as follows:

$$\begin{aligned}(P_x)_{ij}&=&<P_xe_j,e_i>\\&=&\frac{1}{||x||^2}<e_j,x><x,e_i>\\&=&\frac{x_jx_i}{||x||^2}\end{aligned}$$

Thus, we are led to the formula in the statement. □

Regarding now embeddings of $P^{N-1}_{\mathbb R}$ into Euclidean spaces $\mathbb R^n$, many things can be said, with a straightforward construction here being as follows:

Theorem 3.8. *The projective space $P^{N-1}_{\mathbb R}$ is a smooth manifold, with charts*

$$(x_1,\ldots,x_N)\to\left(\frac{x_1}{x_i},\ldots,\frac{x_{i-1}}{x_i},\frac{x_{i+1}}{x_i},\ldots,\frac{x_N}{x_i}\right)$$

where $x_i\neq0$. This manifold is compact, and of dimension $N-1$.

Proof. We know that $P^{N-1}_{\mathbb R}$ appears as the space of lines in $\mathbb R^N$ passing through the origin, so we have the following formula, with $\sim$ being the proportionality of vectors, given as usual by $x\sim y$ when $x=\lambda y$, for some scalar $\lambda\neq0$:

$$P^{N-1}_{\mathbb R}=\mathbb R^N-\{0\}/\sim$$

Alternatively, we can restrict if we want the attention to the vectors on the unit sphere $S^{N-1}_{\mathbb R}\subset\mathbb R^N$, and this because any line in $\mathbb R^N$ passing through the origin will certainly cross this sphere. Moreover, it is clear that our line will cross the sphere in exactly two points $\pm x$, and we conclude that we have the following formula, with $\sim$ being now the proportionality of vectors on the sphere, given by $x\sim y$ when $x=\pm y$:

$$P^{N-1}_{\mathbb R}=S^{N-1}_{\mathbb R}/\sim$$

With this discussion made, let us get now to what is to be proved. Obviously, once we fix an index $i\in\{1,\ldots,N\}$, the condition $x_i\neq0$ on the vectors $x\in\mathbb R^N-\{0\}$ defines an open subset $U_i\subset P^{N-1}_{\mathbb R}$, and the open subsets that we get in this way cover $P^{N-1}_{\mathbb R}$:

$$P^{N-1}_{\mathbb R}=U_1\cup\ldots\cup U_N$$

Moreover, the map in the statement is injective $U_i\to\mathbb R^{N-1}$, and it is clear too that the changes of charts are C^∞. Thus, we have our smooth manifold, as claimed. □

3b. Three dimensions

Let us restrict now the attention to the 3D case. We have here:

DEFINITION 3.9. *The real projective space, denoted $P^3_{\mathbb{R}}$, is the space of lines in $\mathbb{R}^4$ passing through the origin. To be more precise:*

(1) *We call each of the lines in $\mathbb{R}^4$ passing through the origin a point of $P^3_{\mathbb{R}}$.*
(2) *We also call each plane in $\mathbb{R}^4$ passing through the origin a line of $P^3_{\mathbb{R}}$.*

In order to concretely deal now with $P^3_{\mathbb{R}}$, say by using coordinates, as we would prefer, several methods are available. We first have the following result:

THEOREM 3.10. *The projective space $P^3_{\mathbb{R}}$ appears, alternatively, as the quotient*

$$P^3_{\mathbb{R}} = \mathbb{R}^4 - \{0\}/\sim$$

with $\sim$ being the proportionality of vectors, given by $x \sim y$ when $x = \lambda y$, with $\lambda \neq 0$.

PROOF. We know that the projective space $P^3_{\mathbb{R}}$ appears by definition as the space of lines in $\mathbb{R}^4$ passing through the origin, and this gives the result. □

As a continuation of this, we can restrict if we want the attention to the vectors on the unit sphere $S^3_{\mathbb{R}} \subset \mathbb{R}^4$, and this because any line in $\mathbb{R}^4$ passing through the origin will certainly cross this sphere. We are led in this way to the following result:

THEOREM 3.11. *The projective space $P^3_{\mathbb{R}}$ appears also as the quotient*

$$P^3_{\mathbb{R}} = S^3_{\mathbb{R}}/\sim$$

with $\sim$ being the proportionality of vectors on the sphere, given by $x \sim y$ when $x = \pm y$.

PROOF. According to the discussion above, we can restrict the attention to the vectors on the sphere $S^3_{\mathbb{R}} \subset \mathbb{R}^4$, and this gives the following formula, with $\sim$ standing as before for the proportionality of vectors in space, given by $x \sim y$ when $x = \lambda y$, with $\lambda \neq 0$:

$$P^3_{\mathbb{R}} = S^3_{\mathbb{R}}/\sim$$

But, it is clear that our line will cross the sphere in exactly two points $\pm x$, and we conclude that we have the formula in the statement. □

There is some linear algebra to be done here too, and we have:

THEOREM 3.12. *The projective space $P^3_{\mathbb{R}}$ can be thought of as being the space of rank 1 projections in the matrix algebra $M_4(\mathbb{R})$, given by*

$$P_x = \frac{1}{||x||^2}(x_i x_j)_{ij}$$

by identifying the lines in $\mathbb{R}^4$ passing through the origin with the corresponding rank 1 projections in $M_4(\mathbb{R})$, in the obvious way.

PROOF. This is indeed something quite self-explanatory. □

Many other things can be said about $P^3_{\mathbb{R}}$, as a continuation of this.

3c. Curves, surfaces

Curves, surfaces.

3d. Higher dimensions

Higher dimensions.

3e. Exercises

Exercises:

Exercise 3.13.

Exercise 3.14.

Exercise 3.15.

Exercise 3.16.

Exercise 3.17.

Exercise 3.18.

Exercise 3.19.

Exercise 3.20.

Bonus exercise.

CHAPTER 4

Generalizations

4a. Arbitrary fields

We discuss here some generalizations of the theory that we have, the idea being that we can talk about projective spaces over arbitrary fields, in the obvious way.

Let us start with some field theory preliminaries. We first have:

Definition 4.1. *A field is a set F with a sum operation $+$ and a product operation $\times$, subject to the following conditions:*

1. *$a+b=b+a$, $a+(b+c)=(a+b)+c$, there exists $0\in F$ such that $a+0=0$, and any $a\in F$ has an inverse $-a\in F$, satisfying $a+(-a)=0$.*
2. *$ab=ba$, $a(bc)=(ab)c$, there exists $1\in F$ such that $a1=a$, and any $a\neq 0$ has a multiplicative inverse $a^{-1}\in F$, satisfying $aa^{-1}=1$.*
3. *The sum and product are compatible via $a(b+c)=ab+ac$.*

Apparently, the simplest possible field is $\mathbb{Q}$. However, this is not exactly true, because, by a strange twist of fate, the numbers $0,1$, whose presence in a field is mandatory, $0,1\in F$, can form themselves a field, with addition as follows:

$$1+1=0$$

To be more precise, according to our field axioms, we certainly must have:

$$0+0=0\times 0=0\times 1=1\times 0=0$$

$$0+1=1+0=1\times 1=1$$

Thus, everything regarding the addition and multiplication of $0,1$ is uniquely determined, except for the value of $1+1$. And here, you would say that we should normally set $1+1=2$, with $2\neq 0$ being a new field element, but the point is that $1+1=0$ is something natural too, this being the addition modulo 2:

$$1+1=0(2)$$

And, what we get in this way is a field, denoted as follows:

$$\mathbb{F}_2=\{0,1\}$$

Let us summarize this finding, along with a bit more, obtained by suitably replacing our 2, used for addition, with an arbitrary prime number p, as follows:

Theorem 4.2. *The following happen:*

1. $\mathbb{Q}$ *is the simplest field having the property* $1+\ldots+1\neq 0$*, in the sense that any field* F *having this property must contain it,* $\mathbb{Q}\subset F$*.*
2. *The property* $1+\ldots+1\neq 0$ *can hold or not, and if not, the smallest number of terms needed for having* $1+\ldots+1=0$ *is a certain prime number* p*.*
3. $\mathbb{F}_p=\{0,1,\ldots,p-1\}$*, with* p *prime, is the simplest field having the property* $1+\ldots+1=0$*, with* p *terms, in the sense that this implies* $\mathbb{F}_p\subset F$*.*

Proof. All this is basic number theory, the idea being as follows:

(1) This is clear, because $1+\ldots+1\neq 0$ tells us that we have an embedding $\mathbb{N}\subset F$, and then by taking inverses with respect to $+$ and $\times$ we obtain $\mathbb{Q}\subset F$.

(2) Again, this is clear, because assuming $1+\ldots+1=0$, with $p=ab$ terms, chosen minimal, we would have a formula as follows, which is a contradiction:

$$(\underbrace{1+\ldots+1}_{a\ terms})(\underbrace{1+\ldots+1}_{b\ terms})=0$$

(3) This follows a bit as in (1), with the copy $\mathbb{F}_p\subset F$ consisting by definition of the various sums of type $1+\ldots+1$, which must cycle modulo p, as shown by (2). □

Getting back now to our philosophical discussion regarding numbers, what we have in Theorem 4.2 is not exactly good news, suggesting that, on purely mathematical grounds, there is a certain rivalry between $\mathbb{Q}$ and $\mathbb{F}_p$, as being the simplest field.

So, which of these two fields shall we study here, say as having been created first? Not an easy question, and as an answer to this, we have:

Answer 4.3. *Ignoring what pure mathematics might say, and trusting instead physics and chemistry, we will choose to trust in* $\mathbb{Q}$*, as being the simplest field.*

In short, welcome to science, and with this being something quite natural for us, mathematics and science being the topic of the present book.

Moving ahead with some more arithmetic, inside $\mathbb{Q}$ and perhaps other fields too, let us start with the following key theorem of Fermat, for the usual integers:

Theorem 4.4. *We have the following congruence, for any prime* p*,*

$$a^p=a(p)$$

called Fermat's little theorem.

PROOF. The simplest way is to do this by recurrence on $a\in\mathbb N$, as follows:

$$\begin{aligned}(a+1)^p&=\sum_{k=0}^p\binom{p}{k}a^k\\&=a^p+1(p)\\&=a+1(p)\end{aligned}$$

Here we have used the fact that all non-trivial binomial coefficients $\binom{p}{k}$ are multiples of p, as shown by a close inspection of these binomial coeffients, given by:

$$\binom{p}{k}=\frac{p(p-1)\ldots(p-k+1)}{k!}$$

Thus, we have the result for any $a\in\mathbb N$, and with the case $p=2$ being trivial, we can assume $p\geq 3$, and here by using $a\to -a$ we get it for any $a\in\mathbb Z$, as desired. □

The Fermat theorem is particularly interesting when extended from the integers to the arbitrary field case. In order to discuss this question, let us start with:

THEOREM 4.5. *Given a field F, define its characteristic $p=char(F)$ as being the smallest $p\in\mathbb N$ such that the following happens, and as $p=0$, if this never happens:*

$$\underbrace{1+\ldots+1}_{p\ times}=0$$

Then, assuming $p>0$, this characteristic p must be a prime number, we have a field embedding $\mathbb F_p\subset F$, and $q=|F|$ must be of the form $q=p^k$, with $k\in\mathbb N$.

PROOF. Very crowded statement that we have here, the idea being as follows:

(1) The fact that $p>0$ must be prime comes by contradiction, by using:

$$(\underbrace{1+\ldots+1}_{a\ times})\times(\underbrace{1+\ldots+1}_{b\ times})=\underbrace{1+\ldots+1}_{ab\ times}$$

Indeed, assuming that we have $p=ab$ with $a,b>1$, the above formula corresponds to an equality of type $AB=0$ with $A,B\neq 0$ inside F, which is impossible.

(2) Back to the general case, F has a smallest subfield $E\subset F$, called prime field, consisting of the various sums $1+\ldots+1$, and their quotients. In the case $p=0$ we obviously have $E=\mathbb Q$. In the case $p>0$ now, the multiplication formula in (1) shows that the set $S=\{1+\ldots+1\}$ is stable under taking quotients, and so $E=S$.

(3) Now with $E=S$ in hand, we obviously have $(E,+)=\mathbb Z_p$, and since the multiplication is given by the formula in (1), we conclude that we have $E=\mathbb F_p$, as a field. Thus, in the case $p>0$, we have constructed an embedding $\mathbb F_p\subset F$, as claimed.

(4) In the context of the above embedding $\mathbb F_p\subset F$, we can say that F is a vector space over $\mathbb F_p$, and so we have $|F|=p^k$, with $k\in\mathbb N$ being the dimension of this space. □

In relation with Fermat, we can extend the trick in the proof there, as follows:

Proposition 4.6. *In a field F of characteristic $p>0$ we have*

$$(a+b)^p=a^p+b^p$$

for any two elements $a,b\in F$.

Proof. We have indeed the computation, exactly as in the proof of Fermat, by using the fact that the non-trivial binomial coefficients are all multiples of p:

$$(a+b)^p=\sum_{k=0}^p\binom{p}{k}a^kb^{p-k}=a^p+b^p$$

Thus, we are led to the conclusion in the statement. □

Observe that we can iterate the Fermat formula, and we obtain $(a+b)^r=a^r+b^r$ for any power $r=p^s$. In particular we have, with $q=|F|$, the following formula:

$$(a+b)^q=a^q+b^q$$

But this is something quite interesting, showing that the following subset of F, which is closed under multiplication, is closed under addition too, and so is a subfield:

$$E=\left\{a\in F\Big|a^q=a\right\}$$

So, what is this subfield $E\subset F$? In the lack of examples, or general theory for subfields $E\subset F$, we are a bit in the dark here, but it seems quite reasonable to conjecture that we have $E=F$. Thus, our conjecture would be that we have the following formula, for any $a\in F$, and with this being the field extension of the Fermat theorem itself:

$$a^q=a$$

Now that we have our conjecture, let us think at a potential proof. And here, by looking at the proof of the Fermat theorem, the recurrence method from there, based on $a\to a+1$, cannot work as such, and must be suitably fine-tuned.

Thinking a bit, the recurrence from the proof of Fermat somehow rests on the fact that the additive group $\mathbb{Z}$ is singly generated, by $1\in\mathbb{Z}$. Thus, we need some sort of field extension of this single generation result, and in the lack of something additive here, the following theorem, which is something multiplicative, comes to the rescue:

Theorem 4.7. *Given a field F, any finite subgroup of its multiplicative group*

$$G\subset F-\{0\}$$

must be cyclic.

Proof. This can be done via some standard arithmetics, as follows:

(1) Let us pick an element $g\in G$ of highest order, $n=ord(g)$. Our claim, which will easily prove the result, is that the order $m=ord(h)$ of any $h\in G$ satisfies $m|n$.

(2) In order to prove this claim, let $d=(m,n)$, write $d=am+bn$ with $a,b\in\mathbb Z$, and set $k=g^ah^b$. We have then the following computations:

$$k^m=g^{am}h^{bm}=g^{am}=g^{d-bn}=g^d$$

$$k^n=g^{an}h^{bn}=h^{bn}=h^{d-am}=h^d$$

By using either of these formulae, say the first one, we obtain:

$$k^{[m,n]}=k^{mn/d}=(k^m)^{n/d}=(g^d)^{n/d}=g^n=1$$

Thus $ord(k)|[m,n]$, and our claim is that we have in fact $ord(k)=[m,n]$.

(3) In order to prove this latter claim, assume first that we are in the case $d=1$. But here the result is clear, because the formulae in (2) read $g=k^m,h=g^n$, and since $n=ord(g),m=ord(g)$ are prime to each other, we conclude that we have $ord(k)=mn$, as desired. As for the general case, where d is arbitrary, this follows from this.

(4) Summarizing, we have proved our claim in (2). Now since the order $n=ord(g)$ was assumed to be maximal, we must have $[m,n]|n$, and so $m|n$. Thus, we have proved our claim in (1), namely that the order $m=ord(h)$ of any $h\in G$ satisfies $m|n$.

(5) But with this claim in hand, the result follows. Indeed, since the polynomial x^n-1 has all the elements $h\in G$ as roots, its degree must satisfy $n\geq|G|$. On the other hand, from $n=ord(g)$ with $g\in G$, we have $n||G|$. We therefore conclude that we have $n=|G|$, which shows that G is indeed cyclic, generated by the element $g\in G$. □

We can now extend the Fermat theorem to the finite fields, as follows:

Theorem 4.8. *Given a finite field F, with $q=|F|$ we have*

$$a^q=a$$

for any $a\in F$.

Proof. According to Theorem 4.7 the multiplicative group $F-\{0\}$ is cyclic, of order $q-1$. Thus, the following formula is satisfied, for any $a\in F-\{0\}$:

$$a^{q-1}=1$$

Now by multiplying by a, we are led to the conclusion in the statement, with of course the remark that the formula there trivially holds for $a=0$. □

The Fermat polynomial X^p-X is something very useful, and its field generalization X^q-X, with $q=p^k$ prime power, can be used in order to elucidate the structure of finite fields. In order to discuss this question, let us start with a basic fact, as follows:

Proposition 4.9. *Given a finite field F, we have*

$$X^q-X=\prod_{a\in F}(X-a)$$

with $q=|F|$.

Proof. We know from the Fermat theorem above that we have $a^q=a$, for any $a\in F$. We conclude from this that all the elements $a\in F$ are roots of the polynomial X^q-X, and so this polynomial must factorize as in the statement. □

The continuation of the story is more complicated, as follows:

Theorem 4.10. *For any prime power $q=p^k$ there is a unique field $\mathbb{F}_q$ having q elements. At $k=1$ this is the usual $\mathbb{F}_p$, and in general, this is the field making*

$$X^q-X=\prod_{a\in F}(X-a)$$

happen, in some abstract algebraic sense.

Proof. We are punching here a bit above our weight, the idea being as follows:

(1) At $k=1$ there is nothing much to be said, because the prime field embedding $\mathbb{F}_p\subset F$ found in Theorem 4.2 must be an isomorphism. Thus, done with this.

(2) At $k\geq 2$ however, both the construction and uniqueness of $\mathbb{F}_q$ are non-trivial. However, the idea is not that complicated. Indeed, instead of struggling first with finding a model for $\mathbb{F}_q$, and then struggling some more with proving the uniqueness, the point is that we can solve both these problems, at the same time, by looking at X^q-X.

(3) To be more precise, this polynomial X^q-X must have some sort of abstract, minimal "splitting field", and this is how $\mathbb{F}_q$ comes, both existence and uniqueness. We will be back to this, which is something non-trivial, later in this book, with details. □

4b. Discrete geometry

Getting now to geometry over finite fields, we have here the following result:

Theorem 4.11. *Given a field F, we can talk about the projective plane P^2_F, as being the space of lines in F^3 passing through the origin, having cardinality*

$$|P^2_F|=q^2+q+1$$

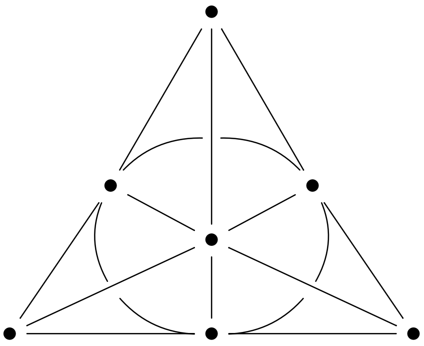
where $q=|F|$, in the case where our field F is finite.

Proof. This is indeed clear from definitions, with the cardinality coming from:

$$|P^2_F|=\frac{|F^3-\{0\}|}{|F-\{0\}|}=\frac{q^3-1}{q-1}=q^2+q+1$$

Thus, we are led to the conclusions in the statement. □

As an example, let us see what happens for the simplest finite field that we know, namely $F=\mathbb{F}_2$. Here our projective plane, having $4+2+1=7$ points, and 7 lines, is a famous combinatorial object, called Fano plane, which is depicted as follows:



Here the circle in the middle is by definition a line, and with this convention, the basic axioms for projective geometry are satisfied, in the sense that any two points determine a line, and any two lines determine a point. And isn't this beautiful.

4c. Complex numbers

Getting now the complex setting, we have here, exactly as in the real case:

Definition 4.12. *We can define the complex projective space $P^{N-1}_\mathbb{C}$ as being the space of complex lines in $\mathbb{C}^N$ passing through the origin.*

As an alternative definition, based this time on linear algebra, we have:

Theorem 4.13. *The complex projective space $P^{N-1}_\mathbb{C}$ is the space of rank 1 projections in the matrix algebra $M_N(\mathbb{C})$, given by*

$$P_x=\frac{1}{||x||^2}(x_i\bar{x}_j)_{ij}$$

by identifying the lines in $\mathbb{C}^N$ passing through the origin with the corresponding rank 1 projections in $M_N(\mathbb{C})$, in the obvious way.

Proof. All this follows indeed via the same arguments as in the real case. □

Talking now differential geometry, the complex projective space $P^{N-1}_\mathbb{C}$ is a smooth compact manifold, having complex dimension $N-1$.

4d. Complex geometry

Complex geometry.

4e. Exercises

Exercises:

Exercise 4.14.

Exercise 4.15.

Exercise 4.16.

Exercise 4.17.

Exercise 4.18.

Exercise 4.19.

Exercise 4.20.

Exercise 4.21.

Bonus exercise.

Part II

Projective manifolds

That's why I go for that rock and roll music
Any old way you choose it
It's got a back beat, you can't lose it
Any old time you use it

CHAPTER 5

Bézout theorem

5a. Plane curves

Plane curves.

5b. Intersections

Intersections.

5c. Bézout theorem

Bézout theorem.

5d. Some applications

Some applications.

5e. Exercises

Exercises:

Exercise 5.1.

Exercise 5.2.

Exercise 5.3.

Exercise 5.4.

Exercise 5.5.

Exercise 5.6.

Exercise 5.7.

Exercise 5.8.

Bonus exercise.

CHAPTER 6

Abstract algebra

6a. Abstract algebra

Let us get now to $\mathbb{R}^3$. Here we are right away into a dillema, because the plane curves have two possible generalizations. First we have the algebraic curves in $\mathbb{R}^3$:

DEFINITION 6.1. *An algebraic curve in $\mathbb{R}^3$ is a curve as follows,*

$$C=\left\{(x,y,z)\in\mathbb{R}^3\Big|P(x,y,z)=0,\ Q(x,y,z)=0\right\}$$

appearing as the joint zeroes of two polynomials P,Q.

These curves look of course like the usual plane curves, and at the level of the phenomena that can appear, these are similar to those in the plane, involving singularities and so on, but also knotting, which is a new phenomenon. However, it is hard to say something with bare hands about knots. We will be back to this, later in this book.

On the other hand, as another natural generalization of the plane curves, and this might sound a bit surprising, we have the surfaces in $\mathbb{R}^3$, constructed as follows:

DEFINITION 6.2. *An algebraic surface in $\mathbb{R}^3$ is a surface as follows,*

$$S=\left\{(x,y,z)\in\mathbb{R}^3\Big|P(x,y,z)=0\right\}$$

appearing as the zeroes of a polynomial P.

The point indeed is that, as it was the case with the plane curves, what we have here is something defined by a single equation. And with respect to many questions, having a single equation matters a lot, and this is why surfaces in $\mathbb{R}^3$ are "simpler" than curves in $\mathbb{R}^3$. In fact, believe me, they are even the correct generalization of the curves in $\mathbb{R}^2$.

As an example of what can be done with surfaces, which is very similar to what we did with the conics $C\subset\mathbb{R}^2$ before, we have the following result:

THEOREM 6.3. *The degree 2 surfaces $S\subset\mathbb{R}^3$, called quadrics, are the ellipsoid*

$$\left(\frac{x}{a}\right)^2+\left(\frac{y}{b}\right)^2+\left(\frac{z}{c}\right)^2=1$$

which is the only compact one, plus 16 more, which can be explicitly listed.

Proof. We will be quite brief here, because we intend to rediscuss all this in a moment, with full details, in arbitrary N dimensions, the idea being as follows:

(1) The equations for a quadric $S\subset\mathbb{R}^2$ are best written as follows, with $A\in M_3(\mathbb{R})$ being a matrix, $B\in M_{1\times 3}(\mathbb{R})$ being a row vector, and $C\in\mathbb{R}$ being a constant:

$$<Au,u>+Bu+C=0$$

(2) By doing now the linear algebra, and we will come back to this in a moment, with details, or by invoking the theorem of Sylvester on quadratic forms, we are left, modulo degeneracy and linear transformations, with signed sums of squares, as follows:

$$\pm x^2\pm y^2\pm z^2=0,1$$

(3) Thus the sphere is the only compact quadric, up to linear transformations, and by applying now linear transformations to it, we are led to the ellipsoids in the statement.

(4) As for the other quadrics, there are many of them, a bit similar to the parabolas and hyperbolas in 2 dimensions, and some work here leads to a 16 item list. □

With this done, instead of further insisting on the surfaces $S\subset\mathbb{R}^3$, or getting into their rivals, the curves $C\subset\mathbb{R}^3$, which appear as intersections of such surfaces, $C=S\cap S'$, let us get instead to arbitrary N dimensions, see what the axiomatics looks like there, with the hope that this will clarify our dimensionality dillema, curves vs surfaces.

So, moving to N dimensions, we have here the following definition, to start with:

Definition 6.4. *An algebraic hypersurface in $\mathbb{R}^N$ is a space of the form*

$$S=\left\{(x_1,\ldots,x_N)\in\mathbb{R}^N\Big|P(x_1,\ldots,x_N)=0,\forall i\right\}$$

appearing as the zeroes of a polynomial $P\in\mathbb{R}[x_1,\ldots,x_N]$.

Again, this is a quite general definition, covering both the plane curves $C\subset\mathbb{R}$ and the surfaces $S\subset\mathbb{R}^2$, which is certainly worth a systematic exploration. But, no hurry with this, for the moment we are here for talking definitons and axiomatics.

In order to have now a full collection of beasts, in all possible dimensions $N\in\mathbb{N}$, and of all possible dimensions $k\in\mathbb{N}$, we must intersect such algebraic hypersurfaces. We are led in this way to the zeroes of families of polynomials, as follows:

Definition 6.5. *An algebraic manifold in $\mathbb{R}^N$ is a space of the form*

$$X=\left\{(x_1,\ldots,x_N)\in\mathbb{R}^N\Big|P_i(x_1,\ldots,x_N)=0,\forall i\right\}$$

with $P_i\in\mathbb{R}[x_1,\ldots,x_N]$ being a family of polynomials.

As a first observation, as already mentioned, such a manifold appears as an intersection of hypersurfaces S_i, those associated to the various polynomials P_i:

$$X = S_1 \cap \ldots \cap S_r$$

There is actually a bit of a discussion needed here, regarding the parameter $r \in \mathbb{N}$, shall we allow this parameter to be $r = \infty$ too, or not. We will discuss this later, with some algebra helping, the idea being that allowing $r = \infty$ forces in fact $r < \infty$.

As an announcement now, good news, what we have in Definition 6.5 is the good and final notion of algebraic manifold, very general, and with the branch of mathematics studying such manifolds being called algebraic geometry. In what follows we will discuss a bit what can be done with this, as a continuation of our previous work on the plane curves, at the elementary level. All this will lead us into the conclusion that we must first develop commutative algebra, and come back to algebraic geometry afterwards.

Let us first look more in detail at the hypersurfaces. We have here:

Theorem 6.6. *The degree 2 hypersurfaces $S \subset \mathbb{R}^N$, called quadrics, are up to degeneracy and to linear transformations the hypersurfaces of the following form,*

$$\pm x_1^2 \pm \ldots \pm x_N^2 = 0, 1$$

and with the sphere being the only compact one.

Proof. We have two statements here, the idea being as follows:

(1) The equations for a quadric $S \subset \mathbb{R}^N$ are best written as follows, with $A \in M_N(\mathbb{R})$ being a matrix, $B \in M_{1 \times N}(\mathbb{R})$ being a row vector, and $C \in \mathbb{R}$ being a constant:

$$< Ax, x > +Bx + C = 0$$

(2) By doing the linear algebra, or by invoking the theorem of Sylvester on quadratic forms, we are left, modulo linear transformations, with signed sums of squares:

$$\pm x_1^2 \pm \ldots \pm x_N^2 = 0, 1$$

(3) To be more precise, with linear algebra, by evenly distributing the terms $x_i x_j$ above and below the diagonal, we can assume that our matrix $A \in M_N(\mathbb{R})$ is symmetric. Thus A must be diagonalizable, and by changing the basis of $\mathbb{R}^N$, as to have it diagonal, our equation becomes as follows, with $D \in M_N(\mathbb{R})$ being now diagonal:

$$< Dx, x > +Ex + F = 0$$

(4) But now, by making squares in the obvious way, which amounts in applying yet another linear transformation to our quadric, the equation takes the following form, with $G \in M_N(-1, 0, 1)$ being diagonal, and with $H \in \{0, 1\}$ being a constant:

$$< Gx, x >= H$$

(5) Now barring the degenerate cases, we can further assume $G \in M_N(-1,1)$, and we are led in this way to the equation claimed in (2) above, namely:

$$\pm x_1^2 \pm \ldots \pm x_N^2 = 0, 1$$

(6) In particular we see that, up to some degenerate cases, namely emptyset and point, the only compact quadric, up to linear transformations, is the one given by:

$$x_1^2 + \ldots + x_N^2 = 1$$

(7) But this is the unit sphere, so are led to the conclusions in the statement. □

Regarding now the examples of hypersurfaces $S \subset \mathbb{R}^N$, or of more general algebraic manifolds $X \subset \mathbb{R}^N$, there are countless of them, and it is impossible to have some discussion started here, without being subjective. The unit sphere $S^{N-1}_\mathbb{R} \subset \mathbb{R}^N$ gets of course the crown from everyone, as being the most important manifold after $\mathbb{R}^N$ itself. But then, passed this sphere, things ramify, depending on what exact applications of algebraic geometry you have in mind. In what concerns me, here is my next favorite example:

Theorem 6.7. *The invertible matrices $A \in M_N(\mathbb{R})$ lie outside the hypersurface*

$$\det A = 0$$

and are therefore dense, in the space of all matrices $M_N(\mathbb{R})$.

Proof. This is something self-explanatory, but with this result being some key in linear algebra, all this is worth a detailed discussion, as follows:

(1) We certainly know from basic linear algebra that a matrix $A \in M_N(\mathbb{R})$ is invertible precisely when it has nonzero determinant, $\det A \neq 0$. Thus, the invertible matrices $A \in M_N(\mathbb{R})$ are located precisely in the complement of the following space:

$$S = \left\{ A \in M_N(\mathbb{R}) \Big| \det A = 0 \right\}$$

(2) We also know from basic linear algebra, or perhaps not so basic linear algebra, that the determinant $\det A$ is a certain polynomial in the entries of A, of degree N:

$$\det \in \mathbb{R}[X_{11}, \ldots, X_{NN}]$$

(3) We conclude from this that the above set S is a degree N algebraic hypersurface in our sense, in the Euclidean space $M_N(\mathbb{R}) \simeq \mathbb{R}^n$, with $n = N^2$.

(4) Now since the complements of non-trivial hypersurfaces $S \subset \mathbb{R}^n$ are obviously dense, and if needing a formal proof here, for our above hypersurface S this is clear, simply by suitably perturbing the matrix, and in general do not worry, we will be back to this, with full details, we are led to the conclusions in the statement. □

As an illustration for the power of our density result, we have:

Theorem 6.8. *Given two matrices $A,B\in M_N(\mathbb{R})$, their products*

$$AB,BA\in M_N(\mathbb{R})$$

have the same characteristic polynomial, $P_{AB}=P_{BA}$.

Proof. This is something quite hard to prove with bare hands, but we can trick by using Theorem 6.7. Indeed, it follows from definitions that the characteristic polynomial of a matrix is invariant under conjugation, in the sense that we have:

$$P_C=P_{ACA^{-1}}$$

Now observe that, when assuming that A is invertible, we have:

$$AB=A(BA)A^{-1}$$

Thus, we obtain the following formula, in the case where A is invertible:

$$P_{AB}=P_{BA}$$

Now by using the density result from Theorem 6.7, we conclude that this formula holds in fact for any matrix A, by continuity, as desired. □

Summarizing, we have some algebraic geometry theory going on, with applications, at least to questions in linear algebra, and presumably in calculus too. Getting back now to the basics, it is in fact possible to do even more generally, as follows:

Definition 6.9. *An algebraic manifold over a field F is a space of the form*

$$X=\left\{(x_1,\ldots,x_N)\in F^N\Big|P_i(x_1,\ldots,x_N)=0,\forall i\right\}$$

with $P_i\in F[x_1,\ldots,x_N]$ being a family of polynomials.

This might seem a bit abstract, but as a first observation, recall that $F=\mathbb{C}$ is a field too, on par with $F=\mathbb{R}$, and even better than it, in certain contexts. For instance quantum mechanics naturally lives over $F=\mathbb{C}$, instead of our usual $F=\mathbb{R}$. Also, in relation with questions in linear algebra, a matrix $A\in M_N(\mathbb{R})$ is much better viewed as matrix $A\in M_N(\mathbb{C})$, because here it has all N eigenvalues, when counted with multiplicities.

In fact, based on this linear algebra observation, and as our first result in complex algebraic geometry, we can improve Theorem 6.8, as follows:

Theorem 6.10. *Given two matrices $A,B\in M_N(\mathbb{C})$, their products*

$$AB,BA\in M_N(\mathbb{C})$$

have the same eigenvalues, with the same multiplicities.

Proof. To start with, Theorem 6.7 holds over $\mathbb{C}$ too, with the invertible matrices $A \in M_N(\mathbb{C})$ being dense, as being complementary to the following hypersurface:

$$\det A = 0$$

But with this in hand, the trick from the proof of Theorem 6.8 applies, and gives:

$$P_{AB} = P_{BA}$$

But this gives the result, because in the complex matrix setting the characteristic polynomial P encodes the eigenvalues, with multiplicities. □

This was for a first result in complex algebraic geometry, perhaps a bit advanced. At the level of more elementary things, the first thought goes to the plane algebraic curves, in a complex sense. But, surprise here, these are the spaces as follows:

$$C = \left\{(x,y) \in \mathbb{C}^2 \Big| P(x,y) = 0\right\}$$

Now when looking at this formula, we realize that our curve $C \subset \mathbb{C}^2$ is in fact something quite complicated, corresponding to a 2-dimensional surface $X \subset \mathbb{R}^4$. But, no worries, we will come back to this regularly. In fact, in what follows, we will be jointly developing our theory over both $F = \mathbb{R}$ and $F = \mathbb{C}$, with such questions in mind.

Getting back now to Definition 6.9 as stated, what about other fields F? Good question, and in answer, I would have a quick exercise for you, as follows:

Exercise 6.11. *Prove that for $n \geq 3$ the following curve,*

$$x^n + y^n = 1$$

has no non-trivial points, $x, y \neq 0$, over $F = \mathbb{Q}$.

Such ideas are very old, going back to the ancient Greeks, and there are many things that can be said about algebraic geometry in its "arithmetic" version, over arbitrary fields F as above. In fact, this is a point where algebraic geometry really shines, with many known advanced results in number theory having been obtained in this way. But more on this later, once we will get more familiar with algebraic geometry over $F = \mathbb{R}, \mathbb{C}$.

6b. Rings and modules

As explained above, in order to better understand our algebraic manifolds, and go beyond what can be done at the elementary level, we are in need of a crash course in abstract algebra in general, and in commutative algebra in particular, with focus on ideals of polynomials. Hang on, many abstract things to follow. But this will be a good investment, useful for topology and for differential geometry too, later in this book.

Let us start with something that we know well, but is worth reminding, namely:

Definition 6.12. *A field is a set F with a sum operation $+$ and a product operation $\times$, subject to the following conditions:*

(1) *$a+b=b+a$, $a+(b+c)=(a+b)+c$, there exists $0\in F$ such that $a+0=0$, and any $a\in F$ has an inverse $-a\in F$, satisfying $a+(-a)=0$.*
(2) *$ab=ba$, $a(bc)=(ab)c$, there exists $1\in F$ such that $a1=a$, and any $a\neq 0$ has an inverse $a^{-1}\in F$, satisfying $aa^{-1}=1$.*
(3) *The sum and product are compatible via $a(b+c)=ab+ac$.*

In other words, a field satisfies what we can normally expect from "numbers", and as basic examples, we have of course $\mathbb{Q},\mathbb{R},\mathbb{C}$. There are many other examples of fields, along the same lines. We can talk for instance about fields like $\mathbb{Q}[\sqrt{2}]$, as follows:

Proposition 6.13. *The following is an intermediate field $\mathbb{Q}\subset F\subset\mathbb{R}$,*

$$\mathbb{Q}[\sqrt{2}]=\left\{a+b\sqrt{2}\Big|a,b\in\mathbb{Q}\right\}$$

and the same happens for any $\mathbb{Q}[\sqrt{n}]$, with $n\neq m^2$ being not a square.

Proof. All the field axioms are clearly satisfied, except perhaps for the inversion axiom. But this axiom is satisfied too, due to the following formula:

$$\frac{1}{a+b\sqrt{2}}=\frac{a-b\sqrt{2}}{a^2-2b^2}$$

Observe that the denominator is indeed nonzero, due to $a^2\neq 2b^2$, which follows by reasoning modulo 2. As for the case of $\mathbb{Q}[\sqrt{n}]$ with $n\neq m^2$, this is similar. □

As another observation now, complementary to this, with our field theory we are not at all away from geometry, quite the opposite. Indeed, while the usual spaces of functions are obviously not fields, geometry and analysis remain around the corner, due to:

Proposition 6.14. *The quotients of complex polynomials, called rational funtions, when written in reduced form, as follows, with P,Q prime to each other,*

$$f=\frac{P}{Q}$$

are well-defined and continuous outside the zeroes $P_f\subset\mathbb{C}$ of Q, called poles of f:

$$f:\mathbb{C}-P_f\to\mathbb{C}$$

Also, these functions are stable under summing, making products and taking inverses,

$$\frac{P}{Q}+\frac{R}{S}=\frac{PS+QR}{QS}\quad,\quad\frac{P}{Q}\cdot\frac{R}{S}=\frac{PR}{QS}\quad,\quad\left(\frac{P}{Q}\right)^{-1}=\frac{Q}{P}$$

so they form a field $\mathbb{C}(X)$, called field of rational functions.

PROOF. Almost everything here is clear from definitions, and with the comment that, in what regards the term "pole", this does not come from the Poles who invented this, but rather from the fact that, when trying to draw the graph of f, or rather imagine that graph, which takes place in $2+2=4$ real dimensions, we are faced with some sort of tent, which is suspended by infinite poles, which lie, guess where, at the poles of f. □

Getting back now to generalities, the simplest example of field appears to be $\mathbb{Q}$. However, this is not exactly true, because the numbers $0,1$, whose presence in a field is mandatory, $0,1\in F$, can form themselves a field, with structure as follows:

$$1+1=0$$

To be more precise, according to our field axioms, all operations of type $a*b$ with $a,b=0,1$ are uniquely determined, except for $1+1$. You would say that we must normally set $1+1=2$, with $2\neq 0$ being a new field element, but the point is that $1+1=0$ is something natural too, this being the addition modulo 2. And, what we get is a field:

$$\mathbb{F}_2=\{0,1\}$$

Let us summarize this finding, along with a bit more, as follows:

PROPOSITION 6.15. *$\mathbb{Q}$ is the simplest field having the property $1+\ldots+1\neq 0$, in the sense that any field F satisfying this condition must contain $\mathbb{Q}$:*

$$\mathbb{Q}\subset F$$

However, in general this fails, for instance for the field $\mathbb{F}_2=\{0,1\}$, with addition $1+1=0$, and more generally for the field $\mathbb{F}_p$ formed by the integers modulo p, with p prime.

PROOF. Here the first assertion is clear, because $1+\ldots+1\neq 0$ tells us that we have an embedding $\mathbb{N}\subset F$, and then by taking inverses with respect to $+$ and $\times$ we obtain $\mathbb{Q}\subset F$. As for the second assertion, this follows from the above discussion. □

As a conclusion, we have now a taste of field theory, with the various examples in Propositions 6.13, 6.14, 6.15 giving us an indication, on what field theory looks like.

Getting back to general theory, now that we have scalars, $\lambda\in F$, let us do some geometry with them. We have here the following straightforward definition:

DEFINITION 6.16. *A vector space V over a field F is a set with a sum operation $+$ and a multiplication by scalars operation $\times$, subject to the following conditions:*

1. *$a+b=b+a$, $a+(b+c)=(a+b)+c$, there exists $0\in V$ such that $a+0=0$, and any $a\in V$ has an inverse $-a\in V$, satisfying $a+(-a)=0$.*
2. *The multiplication by scalars satisfies $(\lambda\mu)a=\lambda(\mu a)$ and $1a=a$, and is compatible with the vector sum via $\lambda(a+b)=\lambda a+\lambda b$.*

Obviously, this is something very familiar, and in practice you can deal with abstract vector spaces as above a bit in the same way as you deal with $\mathbb{R}^N$ or $\mathbb{C}^N$, provided of course that you take some care, in case the field F has the property $1+\ldots+1=0$. Among others, we have the following result, which helps a lot with everything:

Theorem 6.17. *Any finite dimensional vector space V has a basis, and we have*

$$V=F^N$$

with N being the cardinality of the basis, called dimension of V.

Proof. This is something self-explanatory, that you certainly know well in the cases $F=\mathbb{R},\mathbb{C}$, and exercise for you to remember how all that theory was working, and adapt it to the case of arbitrary fields F, with the adaptation being straightforward. □

As an application of this, further building on Proposition 6.15, we have:

Theorem 6.18. *Given a field F, define its characteristic $p=char(F)$ as being the smallest $p\in\mathbb{N}$ such that the following happens, and as $p=0$, if this never happens:*

$$\underbrace{1+\ldots+1}_{p\ times}=0$$

Then, assuming $p>0$, this characteristic p must be a prime number, we have a field embedding $\mathbb{F}_p\subset F$, and $q=|F|$ must be of the form $q=p^k$, with $k\in\mathbb{N}$.

Proof. Quite crowded statement that we have here, the idea being as follows:

(1) The fact that $p>0$ must be prime comes by contradiction, by using:

$$(\underbrace{1+\ldots+1}_{a\ times})\times(\underbrace{1+\ldots+1}_{b\ times})=\underbrace{1+\ldots+1}_{ab\ times}$$

Indeed, assuming that we have $p=ab$ with $a,b>1$, the above formula corresponds to an equality of type $AB=0$ with $A,B\neq 0$ inside F, which is impossible.

(2) Back to the general case, F has a smallest subfield $E\subset F$, called prime field, consisting of the various sums $1+\ldots+1$, and their quotients. In the case $p=0$ we obviously have $E=\mathbb{Q}$. In the case $p>0$ now, the multiplication formula in (1) shows that the set $S=\{1+\ldots+1\}$ is stable under taking quotients, and so $E=S$.

(3) Now with $E=S$ in hand, we obviously have $(E,+)=\mathbb{Z}_p$, and since the multiplication is given by the formula in (1), we conclude that we have $E=\mathbb{F}_p$, as a field. Thus, in the case $p>0$, we have constructed an embedding $\mathbb{F}_p\subset F$, as claimed.

(4) In the context of the above embedding $\mathbb{F}_p\subset F$, we can say that F is a vector space over $\mathbb{F}_p$, and so we have $|F|=p^k$, with $k\in\mathbb{N}$ being the dimension of this space. □

Many other things can be said about fields, and we will be back to this later, when discussing more in detail, following Galois and others, the various characteristic 0 fields that "numbers" can form, and notably the intermediate fields as follows:

$$\mathbb{Q}\subset F\subset\mathbb{C}$$

Moving ahead with more general theory and notions, next in abstract algebra came the rings and ideals, which are more technical objects, defined as follows:

Definition 6.19. *We have notions of rings, modules and ideals, as follows:*

1. *A ring R is a set with operations $+$ and $\times$, satisfying the usual conditions for such operations, except for $ab=ba$, and for $a\neq 0\implies\exists a^{-1}$.*
2. *A module V over a ring R is a vector space, but we will call it ring, and keep the name vector spaces for the modules over fields, $R=F$.*
3. *An ideal $I\subset R$ is a subgroup with the left ideal property $i\in I,r\in R\implies ir\in I$, or the right ideal property $i\in I,r\in R\implies ri\in I$, or both.*

This was a quite crowded statement, but you get the point, with (1) and (2) we are sort of trying to do field and vector space mathematics, over things which are not necessarily fields and vector spaces over them, and (3) is something technical, non-field specific. At the level of examples, these abound, and we have two important ones, as follows:

(1) The integers form a ring, $R=\mathbb{Z}$, which in addition is commutative, $ab=ba$. As obvious module over $\mathbb{Z}$, we have the lattice $V=\mathbb{Z}^N$. Finally, since $R=\mathbb{Z}$ is commutative, the 3 notions of ideals coincide, and these are the subsets $I=a\mathbb{Z}$, with $a\in\mathbb{Z}$.

(2) The matrices over the integers form a ring, $R=M_N(\mathbb{Z})$, which is noncommutative at $N\geq 1$. As obvious module over $M_N(\mathbb{Z})$, we have the lattice $V=\mathbb{Z}^N$. As for the ideals, things here are a bit more complicated, but since at $N=2$ the matrices of type $\left(\begin{smallmatrix}a&b\\0&0\end{smallmatrix}\right)$ form a left ideal which is not a right ideal, and the matrices of type $\left(\begin{smallmatrix}a&0\\b&0\end{smallmatrix}\right)$ form a right ideal which is not a left ideal, at least we know that our 3 types of ideals make sense.

The question that you surely have in mind is, what are ideals good for? Answer:

Proposition 6.20. *For a subgroup $I\subset R$, the following are equivalent:*

1. *I is a two-sided ideal.*
2. *R/I is a ring.*

Proof. This is something which requires some thinking, as follows:

(1) Since the additive group $(R,+)$ is abelian, given an additive subgroup $I\subset R$ we can form the quotient group R/I, which is abelian too, with addition as follows:

$$(a+I)+(b+I)=(a+b+I)$$

Observe that the unit is $(0+I)=I$, and that inverses are given by $(-a+I)$.

(2) The question is now, can we turn this abelian group R/I into a ring? Normally the multiplication can only be as follows, and with this clarifying our statement, with the condition "R/I is a ring" there meaning, with respect to this precise multiplication:

$$(a+I)(b+I)=(ab+I)$$

(3) But, will this work. As a first observation, there is a bit of analogy here with group theory, where $H\subset G$ must be normal in order for G/H to be a group. Thus, our claim is that the ideal condition is somehow the "analogue of normality, in the ring setting".

(4) In practice now, it is quite clear, exactly as in the group theory setting, that everything will be fine, provided that our multiplication is well-defined. And for this multiplication to be well-defined, the following condition must be satisfied:

$$(a+I)=(a'+I)\ ,\ (b+I)=(b'+I)\quad\implies\quad(ab+I)=(a'b'+I)$$

But this amounts in the following condition to be satisfied:

$$a-a'\in I\ ,\ b-b'\in I\quad\implies\quad ab-a'b'\in I$$

(5) Now comes the math. We have the following identity, which shows that if $I\subset R$ is a two-sided ideal, then the above condition is satisfied, and so done:

$$ab-a'b'=a(b-b')+(a-a')b'$$

(6) Conversely now, if the condition in (4) is satisfied, we have in particular:

$$i-0\in I\ ,\ r-r\in I\quad\implies\quad ir-0r\in I$$

$$r-r\in I\ ,\ i-0\in I\quad\implies\quad ri-r0\in I$$

Thus $I\subset R$ must be a two-sided ideal, and this finishes the proof. □

Many things can be said about rings, modules and ideals, and we will be back to this soon. For formulating however a theorem on the subject, we have:

Theorem 6.21. *Assuming that R is commutative and $I\subset R$ is a maximal ideal, in the sense that it is a proper ideal, $I\neq R$, and there is no bigger proper ideal*

$$I\subset J\subset R$$

the quotient ring $F=R/I$ is a field.

Proof. This is something very standard, the idea being as follows:

(1) Before starting, a quick example. We know that over $R=\mathbb{Z}$, the ideals are the subsets $I=p\mathbb{Z}$ with $p\in\mathbb{N}$. But such an ideal is maximal precisely when p is prime, and this is the same as asking for the quotient ring $R/I=\mathbb{Z}_p$ to be a field.

(2) In general now, assume first that R/I is a field. This means that any nonzero element of R/I is invertible, and with our usual conventions for R/I, this reads:

$$\forall a\notin I\ ,\ \exists b\in R\ ,\ (ab+I)=(1+I)$$

Now assume by contradiction that $I \subset R$ is not maximal, so that we have a bigger ideal $I \subset J \subset R$. If we pick $a \in J - I$, we obtain, by the above, the following:

$$a \in J - I \ , \ b \in R \ , \ ab = 1 + i \ , \ i \in I$$

But this is contradictory, because since J is an ideal, containing I, we must have $ab, i \in J$, so we conclude that we have $1 \in J$, and so $J = R$, contradiction.

(3) Conversely, assume now that I is maximal, and assume too, by contradiction, that R/I is not a field. Then we can find a zero divisor in R/I, which reads:

$$(a + I)(b + I) = (I) \ , \ a, b \notin I$$

In other words, we can find $ab \in I$ with $a, b \notin I$. But then, let us look at:

$$I \subset I + aR \subset R$$

(4) What we have in the middle is an ideal, and it is also clear, from $a \notin I$, that the inclusion on the left is proper. As for the inclusion on the right, our claim is that this is proper too. Indeed, assuming otherwise, we would have a formula as follows:

$$i + ac = 1 \ , \ i \in I$$

Now by multiplying everything by b, we obtain from this:

$$ib + acb = b \ , \ i \in I$$

But this is contradictory, because on the left we have $ib \in I$ and $acb = (ab)c \in I$, which gives $b \in I$, contradicting the condition $b \notin I$. Thus, our claim is proved.

(5) But this is the end of the story, because what we just proved is that what we have in (3) is indeed a proper ideal, contradicting the maximality of I, as desired. □

As an interesting application of this, in relation with Theorem 6.18, we have:

Theorem 6.22. *For any prime power $q = p^k$, we can construct a field $\mathbb{F}_q$ having q elements, as being the quotient field*

$$\mathbb{F}_q = \mathbb{F}_p[X]/(Q)$$

of the ring of polynomials $\mathbb{F}_p[X]$ over the integers modulo p, by the ideal generated by an irreducible polynomial $Q \in \mathbb{F}_p[X]$, of degree k.

Proof. There are several things going on here, the idea being as follows:

(1) To start with, given an arbitrary field F, it follows from definitions that the polynomials over it form a ring, $R = F[X]$. Now if we pick any irreducible polynomial $Q \in F[X]$, and denote by $(Q) \subset F[X]$ the ideal generated by this polynomial, this ideal will be maximal, and by Theorem 6.21 the following quotient will be a field:

$$E = F[X]/(Q)$$

(2) Now if we denote by $k \in \mathbb{N}$ the degree of our polynomial Q, it follows from the basic theory of polynomials that we have an isomorphism of vector spaces, as follows:

$$E \simeq F^k$$

(3) Thus, with $F = \mathbb{F}_p$ as field input, we are led to the conclusion in the statement. Of course, there are still a few details to be checked here, with for instance the fact that we have indeed available irreducible polynomials $Q \in \mathbb{F}_p[X]$ of any degree, needing a proof. We will leave this as an exercise, and we will come back to this, with full details, in chapter 3 below. Among others, we will prove there that $\mathbb{F}_q$ does not depend on the choice of $Q \in \mathbb{F}_p[X]$, and in fact is the unique field having $q = p^k$ elements.

(4) Regarding now the best choice of the irreducible polynomial $Q \in \mathbb{F}_p[X]$, providing us with a good model for the finite field $\mathbb{F}_q$, that we can use in practice, this question depends on the value of $q = p^k$, and many things can be said here. All in all, our models are quite similar to $\mathbb{C} = \mathbb{R}[i]$, with i being a formal number satisfying $i^2 = -1$.

(5) To be more precise, at the simplest exponent, $q = 4$, to start with, we can use $Q = X^2 + X + 1$, with this being actually the unique possible choice of a degree 2 irreducible polynomial $Q \in \mathbb{F}_2[X]$, and this leads to a model as follows:

$$\mathbb{F}_4 = \left\{0, 1, a, a+1 \,\middle|\, a^2 = a + 1\right\}$$

(6) Next, at exponents of type $q = p^2$ with $p \geq 3$ prime, we can use $Q = X^2 - r$, with r being a non-square modulo p, and with $(p-1)/2$ choices here. We are led to:

$$\mathbb{F}_{p^2} = \left\{a + b\gamma \,\middle|\, \gamma^2 = r\right\}$$

Here, as before with $\mathbb{F}_4$, our formula is something self-explanatory. Observe the analogy with $\mathbb{C} = \mathbb{R}[i]$, with i being a formal number satisfying $i^2 = -1$. Finally, at $q = p^k$ with $k \geq 3$ things become more complicated, but the main idea remains the same. □

The above result is quite interesting, among others bringing us back to polynomials, and algebraic geometry. In fact, the ring $R = F[X]$ that we used is a particular case of the following types of rings, that we precisely need in algebraic geometry:

$$R = F[X_1, \ldots, X_N]$$

In view of this, I am sure that you have the following question in mind, why having not talked about such polynomial rings right after Definition 6.19, as being the main examples of rings, at least from our algebraic geometry perspective.

Good point, and in answer, we have kept the best for the end. In abstract algebra we have as well a notion of "algebra", and no wonder here, in view of the name, this must be something important. And this notion, generalizing the polynomials, is as follows:

Definition 6.23. *An algebra A over a field F is a ring which is at the same time a vector space, or perhaps vice versa. That is, we have operations $+, \times$ as follows:*

1. *$a+b=b+a$, $a+(b+c)=(a+b)+c$, there exists $0 \in A$ such that $a+0=0$, and any $a \in A$ has an inverse $-a \in A$, satisfying $a+(-a)=0$.*
2. *$a(bc)=(ab)c$, $a(b+c)=ab+ac$, $(a+b)c=ac+bc$ for any $a,b,c \in A$, and $(\lambda\mu)a=\lambda(\mu a)$ for any $\lambda,\mu \in F$ and $a \in A$, and also $1a=a1=a$.*

Quite complicated, you would say, but putting all the axioms for the rings and vector spaces together can only lead to such a crowded definition. In practice, however, this turns to be something quite simple, because all the above axioms are meant to help with our mathematics, by being a sort of "best of" the possible abstract algebra axioms.

But, let us discuss the examples first. And here, we have many of them, with all being related to geometry or analysis of some sort, as follows:

(1) First we have algebra of polynomials $A=F[X]$. This is a very basic algebra, important to us, and with the extra feature that it is commutative, $PQ=QP$.

(2) More generally, we have the algebra of polynomials $A=F[X_1,\ldots,X_N]$. Again, this algebra is very important to us, and is commutative, $PQ=QP$.

(3) Still talking commutative algebras, we have many of them coming from analysis, the general principle being that "functions form algebras". More on this in a moment.

(4) We have as well the algebra of matrices $A=M_N(F)$. Again this is a very basic example, that we know well, which this time is not commutative, $ST \neq TS$.

Obviously, all this is very interesting, and it looks like we hit a big win, with our Definition 2.23. But, no wonder here, algebra can only be about algebras.

Getting now to the algebras of functions, mentioned in (3), we have here the following key result, bringing among others some further light on Theorem 6.21 too:

Theorem 6.24. *Given a compact space X, the following happen:*

1. *The continuous functions $f: X \to \mathbb{C}$ form a complex algebra $C(X)$.*
2. *Given $x \in X$, the functions satisfying $f(x)=0$, form an ideal $I \subset C(X)$.*
3. *This ideal is maximal, and any maximal ideal $I \subset C(X)$ appears in this way.*
4. *In this picture, the fact that the quotient is a field, $C(X)/I=\mathbb{C}$, is clear.*

Proof. All this is self-explanatory, the idea being as follows:

(1) This is clear. Observe that our algebra is commutative, $fg=gf$.

(2) This is again clear, because $f(x)=0$ implies $(fg)(x)=0$.

(3) This follows from some basic topology, via a suitable open cover for X, and we will leave the clarification of all this as an instructive exercise.

(4) This is clear, because $C(X) \to C(X)/I$ maps $f \to f(x) \in \mathbb{C}$. □

There are many other examples of algebras of functions, along these lines. In fact, we can even trick, and view certain algebras, which are certainly not algebras of functions, as algebras of functions too. As an example, here is a wild physics speculation:

Speculation 6.25. *We can view the matrix algebra $M_2(\mathbb{C})$ as being the algebra of functions on some sort of quantum space M_2, according to the following formula:*

$$M_2(\mathbb{C}) = C(M_2)$$

This quantum space M_2 formally has $|M_2| = 4$ points, and appears as a sort of twist of $\{1,2,3,4\}$. Moreover, we can integrate over M_2, according to the formula

$$\int_{M_2} T = \frac{T_{11} + T_{22}}{2}$$

with the underlying measure being positive and of mass 1.

To be more precise here, let us be crazy, and define M_2 according to the formula $C(M_2) = M_2(\mathbb{C})$, without really knowing what we are doing. Then, we have:

$$|M_2| = \dim_{\mathbb{C}} C(M_2) = \dim_{\mathbb{C}} M_2(\mathbb{C}) = 4$$

Next, since we have $M_2(\mathbb{C}) \simeq \mathbb{C}^4$ as vector spaces, which reads $C(M_2) \simeq C(1,2,3,4)$, this suggests that we should have $M_2 \sim \{1,2,3,4\}$, as some sort of twisting operation. But this can be given a mathematical formulation too, the idea being that at the level of standard bases of $C(M_2) \simeq C(1,2,3,4)$, the multiplication gets twisted as follows:

$$e_{ij}e_{kl} = \delta_{jk}e_{il} \quad \longleftrightarrow \quad e_je_k = \delta_{jk}e_j$$

Finally, in what regards the last assertion, this expresses the standard fact that the normalized trace of 2 × 2 matrices $tr = Tr/2$ is unital and positive, in the sense that:

$$tr(1) = 1 \quad , \quad T \geq 0 \implies tr(T) \geq 0$$

Excited about this? Such things come from quantum mechanics, as developed by Heisenberg, and the above space M_2 can be given a precise mathematical sense, and is the entry point to “noncommutative algebraic geometry”. But more on this later, for the moment, we still have work to do on usual, “commutative” algebraic geometry.

6c. The basis theorem

Let us go back now to our general notion of algebraic manifold, from Definition 6.9. There is an interesting link there with the notion of ideal, coming from:

Proposition 6.26. *Given an arbitrary algebraic manifold, appearing as*

$$X=\left\{(x_1,\ldots,x_N)\in F^N\Big|P_i(x_1,\ldots,x_N)=0,\forall i\right\}$$

with $P_i\in F[x_1,\ldots,x_N]$ being a family of polynomials, the following happen:

(1) *Any linear combination $P=\sum\lambda_iP_i$ vanishes on X.*
(2) *More generally, any combination $P=\sum P_iQ_i$ vanishes on X.*
(3) *Thus, any element $P\in(P_i)$, ideal generated by the P_i, vanishes on X.*

Proof. Here (1), and then (2) too, are both clear from definitions, with the convention in both cases that the sums are finite, and (3) is just an abstract reformulation of (2), because the ideal generated by the polynomials P_i is given by:

$$(P_i)=\left\{\sum P_iQ_i\Big|Q_i\in F[x_1,\ldots,x_N]\right\}$$

Thus, we are led to the conclusions in the statement. □

In view of the above result, we can reformulate our notion of algebraic manifold, in commutative algebra terms, as follows:

Proposition 6.27. *The algebraic manifolds are precisely the sets of the form*

$$X=\left\{x\in F^N\Big|P(x)=0,\forall P\in I\right\}$$

with $I\subset F[x_1,\ldots,x_N]$ being a certain ideal.

Proof. In one sense, this comes from Proposition 6.26, and in the other sense this is trivial, because any ideal I can be written as $I=\{P_i|i\in I\}$, with $P_i=i$. □

The above result is quite interesting, and raises a lot of questions about the ideals $I\subset F[x_1,\ldots,x_N]$, and the manifolds $X\subset F^N$ that they produce. What exactly are the ideals $I\subset F[x_1,\ldots,x_N]$? Is the correspondence $I\to X$ bijective? If not, can we make it bijective, by restricting the attention to a suitable class of ideals I? And so on.

We will answer all these questions in due time. Let us start with something very basic, which can obviously be of great use in algebraic geometry, namely:

Theorem 6.28 (Hilbert basis theorem). *Any ideal of polynomials*

$$I\subset F[x_1,\ldots,x_N]$$

is finitely generated, $I=(P_1,\ldots,P_k)$, for some $P_i\in F[x_1,\ldots,x_N]$.

Proof. This is something quite tricky, the idea being as follows:

(1) Following Emmy Noether, let us call a ring R Noetherian when any ideal $I\subset R$ is finitely generated. Equivalently, any increasing sequence of ideals $I_1\subset I_2\subset\ldots$ must stabilize, in the sense that we must have $I_n=I_{n+1}=\ldots$, for some $n\in\mathbb{N}$.

(2) We want to prove that $F[x_1,\ldots,x_N]$ is Noetherian, and we will do this by recurrence on N. Since $R=F$ is clearly Noetherian, as being a field, we are left with proving the recurrence step. And, for this purpose, we will prove something which is a bit more general, namely that if a ring R is Noetherian, then so is the ring $R[X]$.

(3) We do this by contradiction. So, assume that R is Noetherian, and that $R[X]$ is not Noetherian, so that we have an ideal $I\subset R[X]$ which is not finitely generated.

(4) In order to find a contradiction, let us pick $P_1\in I$ of minimial degree $d_1\in\mathbb N$, then $P_2\in I/(P_1)$ of minimal degree $d_2\in\mathbb N$, then $P_3\in I/(P_1,P_2)$ of minimal degree $d_3\in\mathbb N$, and so on. Since our ideal $I\subset R[X]$ was assumed to be not finitely generated, this procedure will not stop, and we obtain an increasing sequence, as follows:

$$d_1\leq d_2\leq d_3\leq\ldots$$

(5) Now let $a_i\in R$ be the leading coefficient of each P_i, and set:

$$J=(a_1,a_2,\ldots)\subset R$$

Since R was assumed to be Noetherian, we can find $n\in\mathbb N$ such that:

$$J=(a_1,\ldots,a_n)$$

Thus, we have a formula as follows, for certain scalars $\lambda_i\in R$:

$$a_{n+1}=\sum_{i=1}^n\lambda_ia_i$$

(6) With this done, consider the following polynomial:

$$Q=\sum_{i=1}^n\lambda_iX^{d_{n+1}-d_i}P_i$$

This polyomial satisfies then $Q\in(P_1,\ldots,P_n)$, and has the same leading coefficient as $P_{n+1}\notin(P_1,\ldots,P_n)$. Thus, the following polynomial has degree $<d_{n+1}$:

$$P_{n+1}-Q\in I/(P_1,\ldots,P_n)$$

But this is a contradiction, so our assumption in (3) was wrong, which finishes the proof of our theorem, as explained in the steps (1-3). □

Getting back now to algebraic manifolds, Theorem 6.28 tells us that in our original Definition 6.9 we can always assume that the family of polynomials $\{P_i\}$ there is finite. Equivalently, in our reformulation from Proposition 6.27, we can say there at the end that $I\subset F[x_1,\ldots,x_N]$ is finitely generated, with this being true by Theorem 6.28.

However, Theorem 6.28 is best remembered geometrically, as follows:

Theorem 6.29. *Any algebraic manifold $X \subset F^N$ appears as a finite intersection of hypersurfaces*

$$X = \bigcap_i X_i$$

with this intersection being obtained by considering the ideal producing X,

$$I \subset F[x_1, \ldots, x_n]$$

writing $I = (P_1, \ldots, P_n)$, and setting $X_i \subset F^N$ to be the set of zeroes of each P_i.

Proof. This is indeed something self-explanatory, coming from Theorem 6.28. □

Moving ahead now, let us investigate more in detail the correspondence $I \to X$ between ideals $I \subset F[x_1, \ldots, x_N]$ and algebraic manifolds $X \subset F^N$. As a first observation, we have in fact correspondences in both senses, constructed as follows:

Proposition 6.30. *Consider the correspondence $I \to X_I$ given by*

$$X_I = \left\{ x \in F^N \Big| P(x) = 0, \forall P \in I \right\}$$

and consider as well the correspondence $X \to I_X$ given by:

$$I_X = \left\{ P \in F[x_1, \ldots, x_n] \Big| P(x) = 0, \forall x \in X \right\}$$

We have then $X_{I_X} = X$, but in the other sense, $I_{X_I} = I$ fails in general.

Proof. Here the first assertion, namely $X_{I_X} = X$, is clear, and the simplest counterexample to $I_{X_I} = I$ comes from the ideal $I = (x^2)$, in $N = 1$ dimensions. Indeed:

$$I = (x^2) \implies X_I = \{0\} \implies I_{X_I} = (x) \neq I$$

Thus, we are led to the conclusions in the statement. □

Let us have now a closer look at $I_{X_I} \neq I$, based on the above study. We have:

Proposition 6.31. *Given an ideal $I \subset R$, define its radical as being:*

$$\sqrt{I} = \left\{ r \in R \Big| \exists n \in \mathbb{N}, r^n \in I \right\}$$

Then this radical is an ideal, having the following properties:

1. *$I = \pi^{-1}(N)$, with $N \subset R$ being the ideal of nilpotent elements, $r^n = 0$ for some $n \in \mathbb{N}$, and with $\pi : R \to I/R$ being the quotient map.*
2. *$I \subset \sqrt{I}$, $\sqrt{\sqrt{I}} = \sqrt{I}$.*
3. *If $\sqrt{I}$ is finitely generated, then $\sqrt{I}^n \subset I$, for some $n \in \mathbb{N}$.*
4. *If $I, J \subset R$, with R assumed Noetherian, then $\sqrt{I} = \sqrt{J}$ precisely when $I^m \subset J$ and $J^n \subset I$ for some $m, n \in \mathbb{N}$.*

PROOF. This is something elementary, and self-explanatory, as follows:

(1) Here everything, including the fact that $N \subset R$ is indeed an ideal, is clear from definitions. Observe that our formula $I = \pi^{-1}(N)$ proves that I is indeed an ideal.

(2) The assertions there are both clear from definitions.

(3) Again, this is something which is clear from definitions.

(4) This assertion, which makes use of the notion of Noetherian ring, that we met in the proof of Theorem 6.28, follows indeed from (3). □

We can now go back to the correspondences in Proposition 6.30, with the following key addition to the material there:

THEOREM 6.32. *Given an algebraic manifold $X \subset F^N$, its ideal, given by*

$$I_X = \left\{ P \in F[x_1, \ldots, x_n] \Big| P(x) = 0, \forall x \in X \right\}$$

is a radical ideal, in the sense that it satisfies the following condition:

$$I_X = \sqrt{I_X}$$

However, even when restricting the attention to the radical ideals, the correspondence $I \to X$ is still not bijective, in general.

PROOF. This is something elementary, the idea being as follows:

(1) The first assertion is clear from definitions, and we have in fact, more generally, the following formula, which is clear as well from definitions:

$$\sqrt{I} \subset I_{X_I}$$

(2) As for the second assertion, a first counterexemple here comes by assuming that our field F is finite. Indeed, while there are finitely many sets, and so finitely many algebraic manifolds $X \subset F^N$, there are infinitely many radical ideals $I \subset F[X_1, \ldots, x_N]$, for instance one for each irreducible polynomial $P \in F[x_1, \ldots, x_N]$.

(3) As an important observation, the second assertion fails for $F = \mathbb{R}$ too, in $N = 1$ dimensions, the simplest counterexample here being as follows:

$$I = (x^2 + 1) \implies X_I = \emptyset \implies I_{X_I} = \mathbb{R}[X] \neq \sqrt{I}$$

In any case, we are led to the conclusions in the statement. □

The problem is now, what to do? We would certainly love to have $I \to X$ bijective, but this does not look very feasible, at least when F is arbitrary. However, we will see in the next section that when assuming that F is algebrically closed, as is for instance the field of the complex numbers $F = \mathbb{C}$, things drastically change, with $I \to X$ becoming bijective, and with this allowing us to develop a lot of non-trivial algebraic geometry.

6d. Nullstellensatz

Let us first recall that $\mathbb{C}$ is algebrically closed, the result being as follows:

Theorem 6.33. *Any polynomial $P \in \mathbb{C}[X]$ decomposes as*

$$P = c(X - a_1) \ldots (X - a_N)$$

with $c \in \mathbb{C}$ and with $a_1, \ldots, a_N \in \mathbb{C}$.

Proof. The problem is that of proving that our polynomial has at least one root, because afterwards we can proceed by recurrence. We prove this by contradiction. So, assume that P has no roots, and pick a number $z \in \mathbb{C}$ where $|P|$ attains its minimum:

$$|P(z)| = \min_{x \in \mathbb{C}} |P(x)| > 0$$

Since $Q(t) = P(z+t) - P(z)$ is a polynomial which vanishes at $t = 0$, this polynomial must be of the form ct^k + higher terms, with $c \neq 0$, and with $k \geq 1$ being an integer. We obtain from this that, with $t \in \mathbb{C}$ small, we have the following estimate:

$$P(z+t) \simeq P(z) + ct^k$$

Now let us write $t = rw$, with $r > 0$ small, and with $|w| = 1$. Our estimate becomes:

$$P(z+rw) \simeq P(z) + cr^k w^k$$

Now recall that we have assumed $P(z) \neq 0$. We can therefore choose $w \in \mathbb{T}$ such that cw^k points in the opposite direction to that of $P(z)$, and we obtain in this way:

$$|P(z+rw)| \simeq |P(z) + cr^k w^k| = |P(z)|(1 - |c|r^k)$$

Now by choosing $r > 0$ small enough, as for the error in the first estimate to be small, and overcame by the negative quantity $-|c|r^k$, we obtain from this:

$$|P(z+rw)| < |P(z)|$$

But this contradicts our definition of $z \in \mathbb{C}$, as a point where $|P|$ attains its minimum. Thus P has a root, and by recurrence it has N roots, as stated. □

Our aim now will be that of developing algebraic geometry over an arbitrary algebrically closed field F, with the main example in mind being the field of complex numbers $F = \mathbb{C}$. We will see that far more things can be said in this case about the algebra of polynomials $A = F[x_1, \ldots, x_N]$, with respect to what we knew before, when F was arbitrary, and with this in hand, we will develop some basic theory for the algebraic manifolds.

Getting back to the discussion from the previous section, we recall from there that the fundamental question of establishing a bijection between ideals $I \subset F[x_1, \ldots, x_N]$ and algebraic manifolds $X \subset F^N$ basically reduces to the question of deciding whether, for an ideal $I \subset F[x_1, \ldots, x_N]$, the following inclusion is an equality or not:

$$\sqrt{I} \subset I_{X_I}$$

We will see that when F is algebrically closed, this inclusion is indeed an equality, with the result being called Hilbert's Nullstellensatz theorem. Getting started now, let us first establish a weak version of the Nullstellensatz, as follows:

THEOREM 6.34 (Weak Nullstellensatz). *If F is algebrically closed, we have*

$$X_I \neq \emptyset$$

for any proper ideal $I \subset F[x_1, \ldots, x_N]$.

PROOF. This is something quite tricky, the idea being as follows:

(1) As a first observation, we have indeed here a Weak Nullstellensatz, because when assuming that the above-mentioned Nullstellensatz holds, we have:

$$\begin{aligned} X_I = \emptyset &\implies I_{X_I} = F[x_1, \ldots, x_N] \\ &\implies \sqrt{I} = F[x_1, \ldots, x_N] \\ &\implies I = F[x_1, \ldots, x_N] \end{aligned}$$

(2) As a second observation, the assumption that F is algebrically closed is really needed, because otherwise we can come with polynomials of type $P = X^2 + 1$, say when $F = \mathbb{R}$, having no zeroes, and so with ideals of type $I = (P) \in F[X]$, with $X_I = \emptyset$.

(3) As a third and last observation, our assumption that F is algebrically closed tells us that any $P \in F[X]$ has zeroes, and based on this, we want to prove that any $I \subset F[x_1, \ldots, x_N]$ has zeroes, $X_I \neq \emptyset$. Which sounds like a quite plausible claim.

(4) Getting to work now, our precise claim, which will prove our theorem, simply by replacing $I \subset F[x_1, \ldots, x_N]$ with a maximal ideal containing it, is that the maximal ideals $I \subset F[x_1, \ldots, x_N]$ are precisely those of the following form, with $a_1, \ldots, a_N \in F$:

$$I = (x_1 - a_1, \ldots, x_N - a_N)$$

(5) In order to prove this latter claim, let us pick a maximal ideal $I \subset F[x_1, \ldots, x_N]$, and consider the following quotient, that we know to be a field:

$$K = F[x_1, \ldots, x_N]/I$$

Our claim in (4), namely $I = (x_1 - a_1, \ldots, x_N - a_N)$, is then equivalent to:

$$K \simeq F$$

Now since F was assumed to be algebrically closed, proving this amounts in proving that K is algebraic over F. And this is what we will prove, by contradiction.

(6) So, asssume that K is purely transcedental over F. By reordering the variables $x_1, \ldots, x_N$, we can assume that $x_1, \ldots, x_k \in K$ are algebrically independent over F, and that $x_{k+1}, \ldots, x_N \in K$ are algebraic over the following subfield:

$$L = K(x_1, \ldots, x_k) \subset K$$

Observe now that K is finitely generated as a L-module. Our claim, based on this, and which will easily prove the theorem, is that L is finitely generated, as a F-algebra.

(7) In short, we are in need here of some commutative algebra input. Inspired by the above, consider a Noetherian ring R, and an intermediate ring as follows:

$$R \subset S \subset R[x_1, \ldots, x_N]$$

Our claim is that if $R[x_1, \ldots, x_N]$ is finitely generated as S-module, then S is finitely generated as S-algebra. Observe that this will prove indeed our claim in (6).

(8) So, let us prove this. For this purpose, let us pick a family of S-module generators $y_1, \ldots, y_m \in R[x_1, \ldots, x_N]$, and write formulae as follows, with $a_{ij}, b_{ijk} \in S$:

$$x_i = \sum_j a_{ij} y_j \quad , \quad y_i y_j = \sum_k b_{ijk} y_k$$

Now if we set $T =< a_{ij}, b_{ijk} >$, this ring being finitely generated over R, it is Noetherian, and since a submodule of a finitely generated module over a Noetherian ring is finitely generated, with this being something general, and elementary, it follows that S is a finitely generated T-module, and so is a finitely generated R-algebra, as claimed.

(9) With this in hand, let us get back to our proof of the Weak Nullstellensatz. Our claim at the end of (6) is now proved, so let us pick algebra generators $z_1, \ldots, z_l \in K$, and write these generators as quotients of polynomials, as follows:

$$z_i = \frac{P_i}{Q_i}$$

(10) Now observe that given any irreducible polynomial $P \in F[x_1, \ldots, x_k]$, the quotient $1/P$ must be a polynomial in the rational functions z_i, and so P must divide at least one Q_i. Thus, we can only have finitely many irreducible polynomials $P \in F[x_1, \ldots, x_k]$, and with this being wrong at $k \geq 1$, we have reached to a contradiction, as desired. □

Still with me I hope, after all this algebra. We can now formulate a main result, namely the Hilbert Nullstellensatz, in its general form, as follows:

Theorem 6.35 (Nullstellensatz). *If F is algebrically closed, we have*

$$I_{X_I} = \sqrt{I}$$

for any ideal $I \subset F[x_1, \ldots, x_N]$.

Proof. This follows from the Weak Nullstellensatz, as follows:

(1) To start with, let us first recall that we trivially have $\sqrt{I} \subset I_{X_I}$, and also that what we want to prove is stronger than the Weak Nullstellensatz. For more on this, and other comments, we refer to the beginning of the proof of the Weak Nullstellensatz.

(2) In practice, we want to prove that given an ideal $I \subset F[x_1, \ldots, x_N]$, any polynomial $P \in F[x_1, \ldots, x_N]$ vanishing on X_I has the property $P^m \in I$, for some $m \in \mathbb{N}$.

(3) For this purpose, we add 1 dimension, and we consider the following ideal:

$$J=< I,x_{N+1}P(x_1,\ldots,x_N)-1 >$$

Since we have $X_J=\emptyset$, the Weak Nullstellensatz applies, and shows that J is trivial.

(4) In order to best interpret this finding, consider the following algebra:

$$F[x_1,\ldots,x_N][P^{-1}]=F[x_1,\ldots,x_{N+1}]/(x_{N+1}P-1)$$

The triviality of J gives then a formula of the following type, with $f_i\in I$:

$$1=f_0+f_1x_{N+1}+\ldots+f_mx_{N+1}^m$$

(5) Now by multiplying by P^m, we obtain from this the following formula:

$$P^m=P^mf_0+P^{m-1}f_1+\ldots+f_m$$

Thus we have $P^m\in I$, as desired. □

With the Nullstellensatz in hand, we can do many things. Assuming as before that F is algebrically closed, for the remainder of this chapter, let us start with:

Definition 6.36. *Given an algebraic manifold $X\subset F^N$, we define the Zariski topology on it by one of the following equivalent conditions:*

1. *The closed sets are the algebraic submanifolds $Y\subset X$.*
2. *$U_f=\{x\in X|f(x)\neq 0\}$ with $f\in F[x_1,\ldots,x_N]$ is a base of open sets.*

Observe that the Zariski topology is not separated, because any two open sets intersect. Observe also that any descreasing sequence of closed subsets $Y_1\supset Y_2\supset\ldots$ must stablilize, with this coming from the fact that $F[x_1,\ldots,x_N]$ is Noetherian. Many other things can be said here, and we will be back to all this in chapter 7.

Also by using algebra and the Nulstellensatz, we can now investigate the functions on our algebraic manifolds, with a key notion of regularity, as follows:

Definition 6.37. *Let $X\subset F^N$ be an algebraic manifold.*

1. *A function $f:X\to F$ is called regular at $x\in X$ if we can write $f=g/h$, with $g,h\in F[x_1,\ldots,x_N]$, in a neighborhood of x.*
2. *More generally, a function $f:X\to F^M$ is called regular if all its components $f_i:X\to F$ are regular, in the above sense.*
3. *A function $f:X\to Y$, with $Y\subset F^M$ algebraic, is called regular when it appears as the restriction of a regular function $f:X\to F^M$ as above.*

Summarizing, we have a good notion of morphisms for the algebraic manifolds, and by using this, we can say that two manifolds are isomorphic, $X\simeq Y$, when we have a regular bijection between them, in both senses. Many things can be said here, and as a key result on the subject, coming from the Nullstellensatz, we have:

Theorem 6.38. *The algebra of regular functions on a manifold $X\subset F^N$ is*

$$A(X)=F[x_1,\ldots,x_N]/I_X$$

with I_X being as usual the ideal of polynomials $P\in F[x_1,\ldots,x_N]$ vanishing on X.

Proof. This follows indeed from the Nullstellensatz. □

Again, many things can be said here, and we will be back to this in chapter 7.

6e. Exercises

Exercises:

Exercise 6.39.

Exercise 6.40.

Exercise 6.41.

Exercise 6.42.

Exercise 6.43.

Exercise 6.44.

Exercise 6.45.

Exercise 6.46.

Bonus exercise.

CHAPTER 7

Projective manifolds

7a. Projective manifolds

Projective manifolds.

7b. Basic examples

Basic examples.

7c. General theory

General theory.

7d. Further examples

Further examples.

7e. Exercises

Exercises:

Exercise 7.1.

Exercise 7.2.

Exercise 7.3.

Exercise 7.4.

Exercise 7.5.

Exercise 7.6.

Exercise 7.7.

Exercise 7.8.

Bonus exercise.

CHAPTER 8

Elliptic curves

8a. Elliptic curves

We will be interested here in the elliptic curves, which are the smooth, projective algebraic curves of genus 1. Under the assumption $char(F)\neq 2,3$, that we will make in this chapter, the corresponding affine curves are as follows, with $a,b\in F$:

$$y^2=x^3+ax+b$$

To be more precise, the above equation defines indeed an elliptic curve, provided that the curve is non-singular, which amounts in saying that the discriminant of the polynomial on the right is nonzero, with this latter condition being as follows:

$$\Delta=-16(4a^3+27b^2)\neq 0$$

Observe that, when assuming that we are over $F=\mathbb{R}$, in the case $\Delta>0$ the affine curve has two components, and in the case $\Delta<0$ it has one component.

In general, in terms of projective coordinates, the above equation reads:

$$y^2z=x^3+axz^2+bz^3$$

Observe that $z=0$ implies $x=0$, and then any choice of y will do. The projective point $(0,y,0)$ is called distinguished point of our elliptic curve, and is denoted 0.

As a last general remark, it follows from the initial definition of the elliptic curves, as being the smooth, projective algebraic curves of genus 1, that when we are over $F=\mathbb{C}$, these curves correspond to certain embeddings of the torus into the complex projective plane. We will be back to this key fact, later on, with full details.

As a first key result about the elliptic curves, we have:

Theorem 8.1. *Given an elliptic curve X, its points form an abelian group, with operation given generically by*

$$p+q+r=0$$

whenever $p,q,r\in X$ are colinear. That is, in the generic case we set $p+q=-r$, where r is the intersection of the line pq with our curve X.

Proof. This statement is something quite compact, the idea being as follows:

(1) Consider an elliptic curve X, coming from an equation $y^2 = x^3 + ax + b$ as above. Since the curve is symmetric with respect to the x-axis, we can define indeed, generically, a sum operation $p + q = -r$ as above, and the unit for it is $0 = -0$.

(2) The above construction works in the generic case, but in order to have the group law defined all over X, we must take care of the various special situations that can appear, too. And here, in the case $p = q$ we can use the tangent there, in the obvious way, and in the other possible special situations, the formula of $p + q$ is straightforward too.

(3) Thus, we obtain indeed an abelian group, and we will use the same notation, X, for this group and for the elliptic curve itself, with the convention from now on that elliptic curve means elliptic curve as before, given with a distinguished point 0.

(4) In the case there is a subfield $E \subset F$ involved, we denote by X_E the elliptic curve over E, which is a group too. In fact, we have a group embedding $X_E \subset X_F$.

(5) Let us do some computations too, in the generic case. In order to sum two points $p = (x_p, y_p)$ and $q = (x_q, y_q)$, we must intersect X with the line pq, and take the opposite r of that point $-r$. But the line pq is of the form $y = sx + c$, with the slope being:

$$s = \frac{y_p - y_q}{x_p - x_q}$$

Now intersecting this line with the curve leads to the following equation:

$$(sx + c)^2 = x^3 + ax + b$$

On the other hand, since x_p, x_q, x_r must be solutions of this equation, this latter equation must coincide with the following degree 3 equation:

$$(x - x_p)(x - x_q)(x - x_r) = 0$$

Now by looking at the coefficient of x^2, we obtain the following formula:

$$s^2 = x_p + x_q + x_r$$

Thus, with the slope s being as before, we obtain the following formulae:

$$x_r = s^2 - x_p - x_q$$

$$y_r = y_p - s(x_p - x_q)$$

(6) Summarizing, we have the abelian group law on X constructed in two possible ways, and with the group embedding $X_E \subset X_F$ from (4) being fully justified too. □

8b. Rational points

In practice now, we are mostly interested in computing the rational points of the elliptic curves, which amounts in studying the following subgroups:

$$X_\mathbb{Q}\subset X_\mathbb{R}$$

As a key result, regarding these subgroups, we have:

Theorem 8.2 (Mordell-Weil). *Given an elliptic curve X, its subgroup of rational points $X_\mathbb{Q}\subset X_\mathbb{R}$ is finitely generated. Thus, we have a decomposition of type*

$$X_\mathbb{Q}=\mathbb{Z}\oplus\ldots\oplus\mathbb{Z}\oplus\mathbb{Z}_{N_1}\oplus\ldots\oplus\mathbb{Z}_{N_k}$$

with finitely many summands, and with $N_1,\ldots,N_k<\infty$.

Proof. This is something quite tricky, the idea being as follows:

(1) The first step, called weak Mordell-Weil theorem, is that of proving that the quotient abelian group $X_\mathbb{Q}/mX_\mathbb{Q}$ is finite, for any $m\geq 2$:

$$\left|X_\mathbb{Q}/mX_\mathbb{Q}\right|<\infty$$

(2) The second step involves the notion of height of a point $r=(x,y)\in X$, which is defined as follows, by writing $x=p/q$, with p,q prime to each other:

$$h(r)=\log\max(|p|,|q|)$$

The point indeed is that one can prove that there are finitely many rational points of height $h(r)\leq K$, for any $K>0$, and this leads to the result. □

The above result is quite interesting, because it splits the study of rational points into two parts, depending on whether these are free points, or torsion points:

(1) First we have the free rational points of our elliptic curve, corresponding to the torsion-free subgroup of the group $X_\mathbb{Q}$ from Theorem 8.2, namely:

$$\mathbb{Z}\oplus\ldots\oplus\mathbb{Z}\subset X_\mathbb{Q}$$

Many things can be said here, notably with the Birch and Swinnerton-Dyer conjecture, asking for the computation of the number of $\mathbb{Z}$ summands, called rank of X.

(2) Then we have the torsion rational points of our elliptic curve, corresponding to the torsion subgroup of the group $X_\mathbb{Q}$ from Theorem 8.2, namely:

$$\mathbb{Z}_{N_1}\oplus\ldots\oplus\mathbb{Z}_{N_k}\subset X_\mathbb{Q}$$

Here, by a theorem of Mazur, the group on the left can take in fact only 15 possible values, namely $\mathbb{Z}_N$ with $N=1,\ldots,10$ and $N=12$, and $\mathbb{Z}_2\times\mathbb{Z}_{2N}$ with $N=1,2,3,4$.

8c. Hasse principle

We discuss now some wild arithmetic tricks, for dealing with equations over the rationals, and with the rational numbers themselves, based on the notion of p-adic number. The idea is very simple, namely that of completing $\mathbb{Q}$ with respect to a different norm, which privileges the prime number p that we have chosen in advance.

Before that, some motivational talk. The dream in arithmetics, usually concerned with solving equations $f = 0$ over the rationals, is something very simple, namely:

Dream 8.3. *I checked that my equation $f = 0$ has solutions modulo p, for any prime p, so my equation must have solutions over $\mathbb{Q}$.*

As a first observation, the dream holds when f is constant, $f = c$. Indeed, ignoring a bit the differences between integers and rationals, $c = 0(p)$ for any prime p means $c = 0$, so our equation is $c = 0$, having any rational number $x \in \mathbb{Q}$ as solution.

Along the same lines, there are some other examples of very simple equations $f = 0$ for which the dream holds. However, such equations are usually so simple, that we can solve them right away, and so our dream for them is not useful. In general, for more complicated equations, our dream remains wrong, and must be fine-tuned.

As a second piece of motivation, let us talk some analysis too. Everything in analytic number theory comes from the Euler formula for the Riemann series, namely:

$$\sum_{n=1}^\infty \frac{1}{n} = \prod_{p\in P}\left(1-\frac{1}{p}\right)^{-1}$$

But this is again something of "local-global" type, with on the left the global quantity, that is, a usual number, which actually happens to be ∞, in our case, and on the right the "local" versions of this number, with respect to the various primes p.

Summarizing, our dream is something important, both from the algebraic and analytic perspective, and is definitely worth a second look, with the aim of fixing it. We are led in this way to the following update to it, which is a bit more modest:

Hope 8.4. *I checked that my equation $f = 0$ has solutions with respect to any prime p, in a suitable sense, so my equation must have solutions over $\mathbb{Q}$.*

So, this will be our plan for this chapter, doing some mathematics, as for this hope come true. We will see that this can indeed be done, with our vague wording above "with respect to any prime p, in a suitable sense" being replaced by something very precise and mathematical, namely "over the p-adics, for any prime p", and with the statement itself being a deep principle in number theory, called Hasse local-global principle.

Getting to work now, let us further reformulate our dreams and hopes, as follows:

Question 8.5. *What are the p-adic numbers, defined with respect to a chosen prime number p, making the local-global principle work?*

In answer, let us temporarily forget about equations, and the local-global principle, and simply pick a prime number p, and look at the world from the perspective of p. So, imagining that we are p, both me and you, what we see is something as follows:

(1) First, we see all sorts of integers $a \in \mathbb{Z}$. Some appear friendly, namely those of the form $a \in p\mathbb{Z}$, while the others, of the form $a \notin p\mathbb{Z}$, appear bizarre and distant.

(2) Moreover, between friends $a \in p\mathbb{Z}$, those of the form $a \in p^2\mathbb{Z}$ appear particularly close. And among them, $a \in p^3\mathbb{Z}$ are truly very close friends. And so on.

(3) Then, we see all sorts of rationals, $r = a/b$, and again, some are close, some are distant, depending on the exact p^k factor, with $k \in \mathbb{Z}$, appearing inside r.

(4) In particular, the rationals of the form $r = 1/p^k$ with $k >> 0$ appear really frightening. Fortunately they are very far away from us, we can barely see them.

(5) And finally, we can see some irrationals $x \notin \mathbb{Q}$ too, but these being uncountable, it is quite hard to figure out how they look like, and are distributed in space.

Very good, so getting back to Earth now, let us write down a definition, based on what we saw in our Prime Number Experience. By focusing on the integers, and more generally the rationals, and leaving the irrationals for later, we have:

Definition 8.6. *Given p prime, we define the p-adic norm of $r \in \mathbb{Q}$ as being:*

$$|r| = p^{-k} \quad , \quad r = p^k\frac{a}{b} \quad , \quad a, b \neq 0(p)$$

Also, we call the integer $k \in \mathbb{Z}$ the p-adic valuation of r, and denote it $k = v(r)$.

Going ahead now with math, the question is, is our Definition 8.6 correct? That is, is $|r|$ indeed a norm? And here, is depends a bit on your background, with mathematicians being a bit dissapointed, to the point of even choosing to stop calling $|r|$ a norm, but physicists and others being fully happy with it, the result being as follows:

Theorem 8.7. *The p-adic norm $|r| = p^{-k}$ is not exactly a norm, but satisfies the following conditions, which are even better:*

(1) *First axiom: $|x| \geq 0$, with $|x| = 0$ when $x = 0$.*
(2) *Modified second axiom: $|xy| = |x| \cdot |y|$.*
(3) *Strong triangle inequality: $|x + y| \leq \max(|x|, |y|)$.*

Proof. All this follows indeed from some simple arithmetics modulo p:

(1) That axiom clearly holds, with the remark that we forgot to say in Definition 8.6 that $v(0)=\infty$, by definition, because any p^k, no matter how big $k\in\mathbb N$ is, divides 0.

(2) As a first observation, the usual second norm axiom, namely $|\lambda x|=||\lambda||\cdot|x|$, with $||.||$ standing here for the usual absolute value of the numbers, definitely fails, and this because all the p-adic norms $|r|$ are by definition integer powers of p, and an arbitrary $\lambda\in\mathbb Q$ will mess up this. However, we have instead $|xy|=|x|\cdot|y|$, coming from:

$$v(xy)=v(x)v(y)$$

And is this good news or not. After some thinking, this modified second axiom is just as good as the failed usual second axiom, because who cares about arbitrary numbers $\lambda\in\mathbb Q$, not viewed from the perspective of p, I mean. More on this in a moment.

(3) Finally, let us look at sums $x+y$. Over the integers $p^k|x,y$ implies $p^k|x+y$, and with a bit of fractions arithmetic, that we will leave here as an easy exercise, the same holds for rationals, in the sense that we have, in terms of the p-adic valuation:

$$v(x+y)\geq\min(v(x),v(y))$$

Thus the p-adic norm itself, $|r|=p^{-v(r)}$, satisfies the following inequality:

$$|x+y|\leq\max(|x|,|y|)$$

Now, what does this inequality mean, geometrically? Good question, and as a first remark, since this is obviously something stronger than the usual triangle inequality satisfied by the norms, $|x+y|\leq|x|+|y|$, we will call it strong triangle inequality. □

Before going ahead, let us further examine the strong triangle inequality found in the above. This is something new to us, and as a further result on it, we have:

Proposition 8.8. *The strong triangle inequality implies*

$$|x|\neq|y|\implies|x+y|=\max(|x|,|y|)$$

and with this being valid for any modified norm, in the sense of Theorem 8.7.

Proof. This is again something elementary, the idea being as follows:

(1) In what regards the p-adic norm, going back to (3) in the proof of Theorem 8.7, we can add there the observation that, trivially over the integers, and then over the rationals too, with a bit of fraction work, the p-adic valuation satisfies:

$$v(x)\neq v(y)\implies v(x+y)=\min(v(x),v(y))$$

Thus the p-adic norm itself satisfies the condition in the statement.

(2) More generally now, and with this being something quite interesting, our claim is that this phenomenon is valid for any generalized norm in the sense of Theorem 8.7. Indeed, assume that $|x|\geq 0$, with $|x|=0$ when $x=0$, as usual, and that:

$$|xy|=|x|\cdot|y|\quad,\quad |x+y|\leq\max(|x|,|y|)$$

In order to prove our result, assume $|x|>|y|$. We then have, trivially:

$$|x+y|\leq\max(|x|,|y|)=|x|$$

(3) In the other sense now, we have to work a bit. We have the following computation, with at the end the observation that the max cannot be $|y|$, because if that would be the case, the inequality that we would obtain would be $|x|\leq|y|$, contradicting $|x|>|y|$:

$$\begin{aligned}|x|&=|(x+y)-y|\\&\leq\max(|x+y|,|y|)\\&=|x+y|\end{aligned}$$

Thus, we have equality in the estimate in (2), as desired. □

Very nice all this, and getting back now to what we have in Theorem 8.7, namely the modified norm axioms there, we can formulate, as a simple consequence:

PROPOSITION 8.9. *The p-adic norm $|r|=p^{-k}$ is not exactly a norm, but*

$$d(x,y)=|x-y|$$

is a distance. Thus, the rationals $\mathbb{Q}$ become in this way a metric space.

PROOF. With the conditions satisfied by the p-norm $|r|$ in hand, it follows, trivially, that $d(x,y)=|x-y|$ is indeed a distance, making $\mathbb{Q}$ a metric space. □

Now let us turn to irrationals. The quite blurry picture that we saw during our Prime Number Experience, and with the blame at that time being on the uncountability of these beasts, in the lack of something better, can be now explained. Indeed, what we saw were not the "usual" irrationals $x\in\mathbb{R}-\mathbb{Q}$, but rather some irrationals $x\in\mathbb{Q}_p-\mathbb{Q}$ viewed from the perspective of p, constructed according to the following result:

THEOREM 8.10. *By completing $\mathbb{Q}$ with respect to the p-adic distance*

$$d(x,y)=|x-y|$$

we obtain a certain field $\mathbb{Q}_p$, called field of p-adic numbers.

PROOF. This is something very standard, with the passage $\mathbb{Q}\to\mathbb{Q}_p$ being very similar to the passage $\mathbb{Q}\to\mathbb{R}$, that we are very familiar with. In fact, some things get even simpler for p-adics, due to the strong triangle inequality satisfied by the norm. □

What is next? Many things, especially in relation with understanding what the p-adic irrationals $x\in\mathbb{Q}_p-\mathbb{Q}$ really are, concretely speaking. But before that, inspired by the theory of usual numbers, $\mathbb{Z}\subset\mathbb{Q}$, we can introduce the p-adic integers, as follows:

Theorem 8.11. *We can introduce the p-adic integers $\mathbf{Z}_p \subset \mathbb{Q}_p$ as being*

$$\mathbf{Z}_p = \left\{ x \in \mathbb{Q}_p \Big| |x| \leq 1 \right\}$$

not to be confused with $\mathbb{Z}_p$, and this is a ring, appearing as completion of $\mathbb{Z} \subset \mathbf{Z}_p$.

Proof. There are several things going on here, the idea being as follows:

(1) We can certainly introduce a set $\mathbf{Z}_p \subset \mathbb{Q}_p$ by the condition in the statement, and the ring axioms are all clear from the modified norm conditions, from Theorem 8.7, the verifications of the fact that $\mathbf{Z}_p$ is stable under sums and products being as follows:

$$|x|, |y| \leq 1 \implies |x + y| \leq \max(|x|, |y|) \leq 1$$

$$|x|, |y| \leq 1 \implies |xy| = |x| \cdot |y| \leq 1$$

(2) Next, since the valuation of a usual integer $x \in \mathbb{Z}$ satisfies $v(x) \geq 0$, the norm satisfies $|x| \leq 1$, and so we have an inclusion $\mathbb{Z} \subset \mathbf{Z}_p$, as in the statement.

(3) With a bit more work, we can see that $\mathbf{Z}_p$ is closed with respect to the p-adic norm, and also, that is appears as the completion of its subring $\mathbb{Z} \subset \mathbf{Z}_p$. □

With this understood, let us get now to the irrationals, and non-integers, and the p-adic numbers in general, viewed as a whole. Obviously, in order to understand them, we must understand well the Cauchy sequences and convergence in $\mathbb{Q}_p$. But here, many surprises are waiting for us, as for instance the following notorious formula:

Proposition 8.12. *We have the following formula,*

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$$

with respect to the p-adic norm.

Proof. By using $p^n \to 0$, with respect to the p-adic norm, we have:

$$\begin{aligned}
\sum_{k=0}^{n-1} p^k &= \frac{1-p^n}{1-p} \\
&= \frac{1}{1-p} - \frac{p^n}{1-p} \\
&\simeq \frac{1}{1-p} - \frac{0}{1-p} \\
&= \frac{1}{1-p}
\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Quite nice the above formula, we are learning new things here, aren't we, and even more spectacular is its $p=2$ particular case, which reads:

$$\sum_{k=0}^\infty 2^k=-1$$

As a matter of doublecheking, this latter formula can be proved as follows:

$$\begin{aligned}\sum_{k=0}^{n-1}2^k&=&2^n-1\\&\simeq&0-1\\&=&-1\end{aligned}$$

But we will not get scared by this. Moving ahead now with our general program, of understanding the Cauchy sequences and convergence in $\mathbb Q_p$, we have:

Theorem 8.13. *Convergence in $\mathbb Q_p$, and corresponding picture of $\mathbb Q_p$.*

Proof. This follows, as usual, from some elementary arithmetic modulo p, with the conclusion being that the arbitrary p-adic numbers $x\in\mathbb Q_p$ have, after all, a quite intuitive interpretation, when it comes to their decimal, or rather p-adic, expansion. □

Finally, again in the analogy with what we know about numbers, we have:

Theorem 8.14. *The field of p-adic numbers $\mathbb Q_p$ can be further enlarged,*

$$\mathbb Q_p\subset\bar{\mathbb Q}_p$$

into an algebrically closed field $\bar{\mathbb Q}_p$, having many interesting properties.

Proof. This follows indeed by using the general $F\to\bar F$ technology coming from Galois theory, and with this being quite similar to the construction $\mathbb R\to\mathbb C$. □

Getting back now to our original motivations, namely equations for the integers and rationals, and the local-global principle for them, that we are dreaming of, we have:

Theorem 8.15. *Hasse local-global principle, and Hasse-Minkowski theorem.*

Proof. Many things can be said here, especially in continuation of our previous study of elliptic curves. The proofs, however, use a lot of non-trivial algebra. We will present here the main ideas, behind these proofs, with some details missing. □

8d. Further results

Further results.

8e. Exercises

Exercises:

Exercise 8.16.

Exercise 8.17.

Exercise 8.18.

Exercise 8.19.

Exercise 8.20.

Exercise 8.21.

Exercise 8.22.

Exercise 8.23.

Bonus exercise.

Part III

Symmetry groups

Well, it's one for the money
Two for the show
Three to get ready
Now go cat go

CHAPTER 9

Projective rotations

9a. Rotation groups

We discuss in this present Part III of this book the various symmetry groups, continuous or discrete, in the projective geometry setting. To start with, in the present chapter we will discuss the projective rotation groups. Let us start with a very basic result, regarding the usual, affine rotation groups, that you surely know well, namely:

THEOREM 9.1. *We have the following results:*

(1) *The rotations of $\mathbb{R}^N$ form the orthogonal group O_N, which is given by:*

$$O_N=\left\{U\in M_N(\mathbb{R})\Big|U^t=U^{-1}\right\}$$

(2) *The rotations of $\mathbb{C}^N$ form the unitary group U_N, which is given by:*

$$U_N=\left\{U\in M_N(\mathbb{C})\Big|U^*=U^{-1}\right\}$$

In addition, we can restrict the attention to the rotations of the corresponding spheres.

PROOF. This is something that you surely know, the idea being as follows:

(1) We know from linear algebra that a linear map $T:\mathbb{R}^N\to\mathbb{R}^N$, written as $T(x)=Ux$ with $U\in M_N(\mathbb{R})$, is a rotation, in the sense that it preserves the distances and the angles, precisely when the associated matrix U is orthogonal, in the following sense:

$$U^t=U^{-1}$$

Thus, we obtain the result. As for the last assertion, this is clear as well, because an isometry of $\mathbb{R}^N$ is the same as an isometry of the unit sphere $S^{N-1}_\mathbb{R}\subset\mathbb{R}^N$.

(2) We know from linear algebra that a linear map $T:\mathbb{C}^N\to\mathbb{C}^N$, written as $T(x)=Ux$ with $U\in M_N(\mathbb{C})$, is a rotation, in the sense that it preserves the distances and the scalar products, precisely when the associated matrix U is unitary, in the following sense:

$$U^*=U^{-1}$$

Thus, we obtain the result. As for the last assertion, this is clear as well, because an isometry of $\mathbb{C}^N$ is the same as an isometry of the unit sphere $S^{N-1}_\mathbb{C}\subset\mathbb{C}^N$. □

In order to introduce now some further examples of continuous groups $G\subset U_N$, we will need the following standard fact, that you surely know well too:

Proposition 9.2. *We have the following results:*

1. *For an orthogonal matrix $U\in O_N$ we have* $\det U\in\{\pm1\}$.
2. *For a unitary matrix $U\in U_N$ we have* $\det U\in\mathbb T$.

Proof. This is clear from the equations defining O_N,U_N, as follows:

(1) We have indeed the following implications:

$$\begin{aligned}U\in O_N&\implies U^t=U^{-1}\\&\implies\det U^t=\det U^{-1}\\&\implies\det U=(\det U)^{-1}\\&\implies\det U\in\{\pm1\}\end{aligned}$$

(2) We have indeed the following implications:

$$\begin{aligned}U\in U_N&\implies U^*=U^{-1}\\&\implies\det U^*=\det U^{-1}\\&\implies\overline{\det U}=(\det U)^{-1}\\&\implies\det U\in\mathbb T\end{aligned}$$

Here we have used the fact that $\bar z=z^{-1}$ means $z\bar z=1$, and so $z\in\mathbb T$. □

We can now introduce the subgroups $SO_N\subset O_N$ and $SU_N\subset U_N$, as being the subgroups consisting of the rotations which preserve the orientation, as follows:

Theorem 9.3. *The following are groups of matrices,*

$$SO_N=\left\{U\in O_N\Big|\det U=1\right\}\quad,\quad SU_N=\left\{U\in U_N\Big|\det U=1\right\}$$

consisting of the rotations which preserve the orientation.

Proof. The fact that we have indeed groups follows from the properties of the determinant, of from the property of preserving the orientation, which is clear as well. □

Summarizing, we have constructed so far 4 continuous groups of matrices, consisting of various rotations, with inclusions between them, as follows:

$$\begin{tikzcd}SU_N \arrow[r] & U_N\\ SO_N \arrow[u] \arrow[r] & O_N \arrow[u]\end{tikzcd}$$

At $N=1$ the situation is trivial, and we obtain very simple groups, as follows:

Proposition 9.4. *The basic continuous groups at $N = 1$ are*

$$\begin{tikzcd}
\{1\} \arrow[r] & \mathbb{T} \\
\{1\} \arrow[r] \arrow[u] & \{\pm 1\} \arrow[u]
\end{tikzcd}$$

or, equivalently, are the following cyclic groups,

$$\begin{tikzcd}
\mathbb{Z}_1 \arrow[r] & \mathbb{Z}_\infty \\
\mathbb{Z}_1 \arrow[r] \arrow[u] & \mathbb{Z}_2 \arrow[u]
\end{tikzcd}$$

with the convention that $\mathbb{Z}_s$ is the group of s-th roots of unity.

Proof. This is clear from definitions, because for a 1×1 matrix the unitarity condition reads $\bar{U} = U^{-1}$, and so $U \in \mathbb{T}$, and this gives all the results. □

At $N = 2$ now, let us first discuss the real case. The result here is as follows:

Theorem 9.5. *We have the following results:*

(1) *SO_2 is the group of usual rotations in the plane, which are given by:*

$$R_t = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$$

(2) *O_2 consists in addition of the usual symmetries in the plane, given by:*

$$S_t = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}$$

(3) *Abstractly speaking, we have isomorphisms as follows:*

$$SO_2 \simeq \mathbb{T} \quad , \quad O_2 = \mathbb{T} \rtimes \mathbb{Z}_2$$

(4) *When discretizing all this, by replacing the 2-dimensional unit sphere $\mathbb{T}$ by the regular N-gon, the latter isomorphism discretizes as $D_N = \mathbb{Z}_N \rtimes \mathbb{Z}_2$.*

Proof. This follows from some elementary computations, as follows:

(1) The first assertion is clear, because only the rotations of the plane in the usual sense preserve the orientation. As for the formula of R_t, this is something that we know well from linear algebra, obtained by computing $R_t\binom{1}{0}$ and $R_t\binom{0}{1}$.

(2) The first assertion is clear, because rotations left aside, we are left with the symmetries of the plane, in the usual sense. As for formula of S_t, this is something that we know well too, obtained by computing $S_t\binom{1}{0}$ and $S_t\binom{0}{1}$.

(3) The first assertion is clear, because the angles $t \in \mathbb{R}$, taken as usual modulo 2π, form the group $\mathbb{T}$. As for the second assertion, the proof here is similar to the proof of the crossed product decomposition $D_N = \mathbb{Z}_N \rtimes \mathbb{Z}_2$ for the dihedral groups.

(4) This is something more speculative, the idea here being that the isomorphism $O_2 = \mathbb{T} \rtimes \mathbb{Z}_2$ appears from $D_N = \mathbb{Z}_N \rtimes \mathbb{Z}_2$ by taking the $N \to \infty$ limit. □

9b. Pauli matrices

Moving forward, let us keep working out what happens at $N = 2$, but this time with a study in the complex case. We first have here the following key result:

Theorem 9.6. *We have the following formula,*

$$SU_2 = \left\{ \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \ \Big| \ |a|^2 + |b|^2 = 1 \right\}$$

which makes SU_2 isomorphic to the unit sphere $S^1_{\mathbb{C}} \subset \mathbb{C}^2$.

Proof. Consider indeed an arbitrary 2×2 matrix, written as follows:

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Assuming that we have $\det U = 1$, the inverse must be given by:

$$U^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

On the other hand, assuming $U \in U_2$, the inverse must be the adjoint:

$$U^{-1} = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix}$$

We are therefore led to the following equations, for the matrix entries:

$$d = \bar{a} \quad , \quad c = -\bar{b}$$

Thus our matrix must be of the following special form:

$$U = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$$

Moreover, since the determinant is 1, we must have, as stated:

$$|a|^2 + |b|^2 = 1$$

Thus, we are done with one inclusion. As for the converse, this is clear, the matrices in the statement being unitaries, and of determinant 1, and so being elements of SU_2. Finally, regarding the last assertion, recall that the unit sphere $S^1_{\mathbb{C}} \subset \mathbb{C}^2$ is given by:

$$S^1_{\mathbb{C}} = \left\{ (a,b) \ \Big| \ |a|^2 + |b|^2 = 1 \right\}$$

Thus, we have an isomorphism of compact spaces, as follows:

$$SU_2 \simeq S^1_{\mathbb{C}} \quad , \quad \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \to (a,b)$$

We have therefore proved our theorem. □

Regarding now the unitary group U_2, the result here is similar, as follows:

THEOREM 9.7. *We have the following formula,*

$$U_2 = \left\{ d \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \ \Big| \ |a|^2 + |b|^2 = 1, |d| = 1 \right\}$$

which makes U_2 be a quotient compact space, as follows,

$$S^1_{\mathbb{C}} \times \mathbb{T} \to U_2$$

but with this parametrization being no longer bijective.

PROOF. In one sense, this is clear from Theorem 9.6, because we have:

$$|d| = 1 \implies dSU_2 \subset U_2$$

In the other sense, let us pick an arbitrary matrix $U \in U_2$. We have then:

$$\begin{aligned} |\det(U)|^2 &= \det(U)\overline{\det(U)} \\ &= \det(U)\det(U^*) \\ &= \det(UU^*) \\ &= \det(1) \\ &= 1 \end{aligned}$$

Consider now the following complex number, defined up to a sign choice:

$$d = \sqrt{\det U}$$

We know from Proposition 9.2 that we have $|d| = 1$. Thus the rescaled matrix $V = U/d$ is unitary, $V \in U_2$. As for the determinant of this matrix, this is given by:

$$\begin{aligned} \det(V) &= \det(U/d) \\ &= \det(U)/d^2 \\ &= \det(U)/\det(U) \\ &= 1 \end{aligned}$$

Thus we have $V\in SU_2$, and so we can write, with $|a|^2+|b|^2=1$:

$$V=\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}$$

Thus the matrix $U=dV$ appears as in the statement. Finally, observe that the result that we have just proved provides us with a quotient map as follows:

$$S^1_\mathbb{C}\times\mathbb{T}\to U_2\quad,\quad((a,b),d)\to d\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}$$

However, the parametrization is no longer bijective, because when we globally switch signs, the element $((-a,-b),-d)$ produces the same element of U_2. □

Here is now a useful reformulation of our main result so far regarding SU_2, obtained by further building on the parametrization found above:

Theorem 9.8. *We have the formula*

$$SU_2=\left\{\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}\Big|\,x^2+y^2+z^2+t^2=1\right\}$$

which makes SU_2 isomorphic to the unit real sphere $S^3_\mathbb{R}\subset\mathbb{R}^3$.

Proof. We recall from Theorem 9.6 that we have the following formula:

$$SU_2=\left\{\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}\Big|\,|a|^2+|b|^2=1\right\}$$

Now let us write our parameters $a,b\in\mathbb{C}$, which belong to the complex unit sphere $S^1_\mathbb{C}\subset\mathbb{C}^2$, in terms of their real and imaginary parts, as follows:

$$a=x+iy\quad,\quad b=z+it$$

In terms of $x,y,z,t\in\mathbb{R}$, our formula for a generic matrix $U\in SU_2$ becomes the one in the statement. As for the condition to be satisfied by the parameters $x,y,z,t\in\mathbb{R}$, this comes the condition $|a|^2+|b|^2=1$ to be satisfied by $a,b\in\mathbb{C}$, which reads:

$$x^2+y^2+z^2+t^2=1$$

Thus, we are led to the conclusion in the statement. Regarding now the last assertion, recall that the unit sphere $S^3_\mathbb{R}\subset\mathbb{R}^4$ is given by:

$$S^3_\mathbb{R}=\left\{(x,y,z,t)\,\Big|\,x^2+y^2+z^2+t^2=1\right\}$$

Thus, we have an isomorphism of compact spaces, as follows:

$$SU_2\simeq S^3_\mathbb{R}\quad,\quad\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}\to(x,y,z,t)$$

We have therefore proved our theorem. □

As a philosophical comment here, the above parametrization of SU_2 is something very nice, because the parameters (x,y,z,t) range now over the sphere of space-time. Thus, we are probably doing some kind of physics here. More on this later.

Regarding now the group U_2, we have here a similar result, as follows:

THEOREM 9.9. *We have the following formula,*

$$U_2=\left\{(p+iq)\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}\Big|\,x^2+y^2+z^2+t^2=1,\ p^2+q^2=1\right\}$$

which makes U_2 be a quotient compact space, as follows,

$$S^3_\mathbb{R}\times S^1_\mathbb{R}\to U_2$$

but with this parametrization being no longer bijective.

PROOF. We recall from Theorem 9.7 that we have the following formula:

$$U_2=\left\{d\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}\Big|\,|a|^2+|b|^2=1,\ |d|=1\right\}$$

Now let us write our parameters $a,b\in\mathbb{C}$, which belong to the complex unit sphere $S^1_\mathbb{C}\subset\mathbb{C}^2$, and $d\in\mathbb{T}$, in terms of their real and imaginary parts, as follows:

$$a=x+iy\quad,\quad b=z+it\quad,\quad d=p+iq$$

In terms of these new parameters $x,y,z,t,p,q\in\mathbb{R}$, our formula for a generic matrix $U\in SU_2$, that we established before, reads:

$$U=(p+iq)\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}$$

As for the condition to be satisfied by the parameters $x,y,z,t,p,q\in\mathbb{R}$, this comes the conditions $|a|^2+|b|^2=1$ and $|d|=1$ to be satisfied by $a,b,d\in\mathbb{C}$, which read:

$$x^2+y^2+z^2+t^2=1\quad,\quad p^2+q^2=1$$

Thus, we are led to the conclusion in the statement. Regarding now the last assertion, recall that the unit spheres $S^3_\mathbb{R}\subset\mathbb{R}^4$ and $S^1_\mathbb{R}\subset\mathbb{R}^2$ are given by:

$$S^3_\mathbb{R}=\left\{(x,y,z,t)\Big|\,x^2+y^2+z^2+t^2=1\right\}$$

$$S^1_\mathbb{R}=\left\{(p,q)\Big|\,p^2+q^2=1\right\}$$

Thus, we have quotient map of compact spaces, as follows:

$$S^3_\mathbb{R}\times S^1_\mathbb{R}\to U_2\quad,\quad((x,y,z,t),(p,q))\to(p+iq)\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}$$

However, the parametrization is no longer bijective, because when we globally switch signs, the element $((-x,-y,-z,-t),(-p,-q))$ produces the same element of U_2. □

Here is now another reformulation of our main result so far, regarding SU_2, obtained by further building on the parametrization from Theorem 9.8:

Theorem 9.10. *We have the following formula,*

$$SU_2=\left\{xc_1+yc_2+zc_3+tc_4\ \Big|\ x^2+y^2+z^2+t^2=1\right\}$$

where c_1,c_2,c_3,c_4 are matrices given by

$$c_1=\begin{pmatrix}1&0\\0&1\end{pmatrix}\quad,\quad c_2=\begin{pmatrix}i&0\\0&-i\end{pmatrix}$$

$$c_3=\begin{pmatrix}0&1\\-1&0\end{pmatrix}\quad,\quad c_4=\begin{pmatrix}0&i\\i&0\end{pmatrix}$$

called Pauli spin matrices.

Proof. We recall from Theorem 9.8 that the group SU_2 can be parametrized by the real sphere $S^3_\mathbb{R}\subset\mathbb{R}^4$, in the following way:

$$SU_2=\left\{\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}\ \Big|\ x^2+y^2+z^2+t^2=1\right\}$$

Thus, the elements $U\in SU_2$ are precisely the matrices as follows, depending on parameters $x,y,z,t\in\mathbb{R}$ satisfying $x^2+y^2+z^2+t^2=1$:

$$U=x\begin{pmatrix}1&0\\0&1\end{pmatrix}+y\begin{pmatrix}i&0\\0&-i\end{pmatrix}+z\begin{pmatrix}0&1\\-1&0\end{pmatrix}+t\begin{pmatrix}0&i\\i&0\end{pmatrix}$$

But this gives the formula for SU_2 in the statement. □

The above result is often the most convenient one, when dealing with SU_2. This is because the Pauli matrices have a number of remarkable properties, which are very useful when doing computations. These properties can be summarized as follows:

Theorem 9.11. *The Pauli matrices multiply according to the formulae*

$$c_2^2=c_3^2=c_4^2=-1$$

$$c_2c_3=-c_3c_2=c_4$$

$$c_3c_4=-c_4c_3=c_2$$

$$c_4c_2=-c_2c_4=c_3$$

they conjugate according to the following rules,

$$c_1^*=c_1\ ,\ c_2^*=-c_2\ ,\ c_3^*=-c_3\ ,\ c_4^*=-c_4$$

and they form an orthonormal basis of $M_2(\mathbb{C})$, with respect to the scalar product

$$<a,b>=tr(ab^*)$$

with $tr:M_2(\mathbb{C})\to\mathbb{C}$ being the normalized trace of 2×2 matrices, $tr=Tr/2$.

Proof. The first two assertions, regarding the multiplication and conjugation rules for the Pauli matrices, follow from some elementary computations. As for the last assertion, this follows by using these rules. Indeed, the fact that the Pauli matrices are pairwise orthogonal follows from computations of the following type, for $i \neq j$:

$$< c_i, c_j >= tr(c_i c_j^*) = tr(\pm c_i c_j) = tr(\pm c_k) = 0$$

As for the fact that the Pauli matrices have norm 1, this follows from:

$$< c_i, c_i >= tr(c_i c_i^*) = tr(\pm c_i^2) = tr(c_1) = 1$$

Thus, we are led to the conclusion in the statement. □

We should mention here that the Pauli matrices are cult objects in physics, due to the fact that they describe the spin of the electron. Indeed, a bit like our Earth spins around its axis, the electrons spin too. And it took scientists a lot of skill in order to understand the physics and mathematics of the spin, the conclusion being that the Schrödinger wave function space for the electron $H = L^2(\mathbb{R}^3)$ has to be enlarged with a copy of the space $K = \mathbb{C}^2$, via a direct sum, as to take into account the spin, and with this spin being described by the Pauli matrices, in some appropriate, quantum mechanical sense.

As usual, we refer to Feynman [32], Griffiths [42] or Weinberg [95] for more on all this. And with the remark that the Pauli matrices are actually subject to several possible normalizations, depending on formalism, but let us not get into all this here.

9c. Euler-Rodrigues

Back to mathematics, let us discuss now the basic unitary groups in 3 or more dimensions. The situation here becomes fairly complicated, but it is possible however to explicitly compute the rotation groups SO_3 and O_3, and explaining this result, due to Euler-Rodrigues, which is something non-trivial and very useful, will be our next goal.

The proof of the Euler-Rodrigues formula is something quite tricky. Let us start with the following construction, whose usefulness will become clear in a moment:

Proposition 9.12. *The adjoint action $SU_2 \curvearrowright M_2(\mathbb{C})$, given by*

$$T_U(M) = UMU^*$$

leaves invariant the following real vector subspace of $M_2(\mathbb{C})$,

$$E = span_{\mathbb{R}}(c_1, c_2, c_3, c_4)$$

and we obtain in this way a group morphism $SU_2 \to GL_4(\mathbb{R})$.

Proof. We have two assertions to be proved, as follows:

(1) We must first prove that, with $E \subset M_2(\mathbb{C})$ being the real vector space in the statement, we have the following implication:

$$U \in SU_2, M \in E \implies UMU^* \in E$$

But this is clear from the multiplication rules for the Pauli matrices, from Theorem 9.11. Indeed, let us write our matrices U, M as follows:

$$U = xc_1 + yc_2 + zc_3 + tc_4$$

$$M = ac_1 + bc_2 + cc_3 + dc_4$$

We know that the coefficients x, y, z, t and a, b, c, d are real, due to $U \in SU_2$ and $M \in E$. The point now is that when computing UMU^*, by using the various rules from Theorem 9.11, we obtain a matrix of the same type, namely a combination of c_1, c_2, c_3, c_4, with real coefficients. Thus, we have $UMU^* \in E$, as desired.

(2) In order to conclude, let us identify $E \simeq \mathbb{R}^4$, by using the basis c_1, c_2, c_3, c_4. The result found in (1) shows that we have a correspondence as follows:

$$SU_2 \to M_4(\mathbb{R}) \quad , \quad U \to (T_U)_{|E}$$

Now observe that for any $U \in SU_2$ and any $M \in M_2(\mathbb{C})$ we have:

$$T_{U^*}T_U(M) = U^*UMU^*U = M$$

Thus $T_{U^*} = T_U^{-1}$, and so the correspondence that we found can be written as:

$$SU_2 \to GL_4(\mathbb{R}) \quad , \quad U \to (T_U)_{|E}$$

But this a group morphism, due to the following computation:

$$T_UT_V(M) = UVMV^*U^* = T_{UV}(M)$$

Thus, we are led to the conclusion in the statement. □

The point now, which makes the link with SO_3, and which will ultimately elucidate the structure of SO_3, is that Proposition 9.12 can be improved as follows:

Theorem 9.13. *The adjoint action $SU_2 \curvearrowright M_2(\mathbb{C})$, given by*

$$T_U(M) = UMU^*$$

leaves invariant the following real vector subspace of $M_2(\mathbb{C})$,

$$F = span_{\mathbb{R}}(c_2, c_3, c_4)$$

and we obtain in this way a group morphism $SU_2 \to SO_3$.

Proof. We can do this in several steps, as follows:

(1) Our first claim is that the group morphism $SU_2 \to GL_4(\mathbb{R})$ constructed in Proposition 9.12 is in fact a morphism $SU_2 \to O_4$. In order to prove this, recall the following formula, valid for any $U \in SU_2$, from the proof of Proposition 9.12:

$$T_{U^*} = T_U^{-1}$$

We want to prove that the matrices $T_U \in GL_4(\mathbb{R})$ are orthogonal, and in view of the above formula, it is enough to prove that we have:

$$T_U^* = (T_U)^t$$

So, let us prove this. For any two matrices $M, N \in E$, we have:

$$\begin{aligned} < T_{U^*}(M), N > &= < U^*MU, N > \\ &= tr(U^*MUN) \\ &= tr(MUNU^*) \end{aligned}$$

On the other hand, we have as well the following formula:

$$\begin{aligned} < (T_U)^t(M), N > &= < M, T_U(N) > \\ &= < M, UNU^* > \\ &= tr(MUNU^*) \end{aligned}$$

Thus we have indeed $T_U^* = (T_U)^t$, which proves our $SU_2 \to O_4$ claim.

(2) In order now to finish, recall that we have by definition $c_1 = 1$, as a matrix. Thus, the action of SU_2 on the vector $c_1 \in E$ is given by:

$$T_U(c_1) = Uc_1U^* = UU^* = 1 = c_1$$

We conclude that $c_1 \in E$ is invariant under SU_2, and by orthogonality the following subspace of E must be invariant as well under the action of SU_2:

$$e_1^\perp = span_{\mathbb{R}}(c_2, c_3, c_4)$$

Now if we call this subspace F, and we identify $F \simeq \mathbb{R}^3$ by using the basis c_2, c_3, c_4, we obtain by restriction to F a morphism of groups as follows:

$$SU_2 \to O_3$$

But since this morphism is continuous and SU_2 is connected, its image must be connected too. Now since the target group decomposes as $O_3 = SO_3 \sqcup (-SO_3)$, and $1 \in SU_2$ gets mapped to $1 \in SO_3$, the whole image must lie inside SO_3, and we are done. □

The above result is quite interesting, because we will see in a moment that the morphism $SU_2 \to SO_3$ constructed there is surjective. Thus, we will have a way of parametrizing the elements $V \in SO_3$ by elements $U \in SO_2$, and so ultimately by parameters

$(x,y,z,t)\in S^3_{\mathbb{R}}$. In order to work out all this, let us start with the following result, coming as a continuation of Proposition 9.12, independently of Theorem 9.13:

THEOREM 9.14. *With respect to the standard basis c_1,c_2,c_3,c_4 of the vector space $\mathbb{R}^4=span(c_1,c_2,c_3,c_4)$, the morphism $T:SU_2\to GL_4(\mathbb{R})$ is given by:*

$$T_U=\begin{pmatrix}1&0&0&0\\0&x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\0&2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\0&2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2\end{pmatrix}$$

Thus, when looking at T as a group morphism $SU_2\to O_4$, what we have in fact is a group morphism $SU_2\to O_3$, and even $SU_2\to SO_3$.

PROOF. With notations from Proposition 9.12 and its proof, let us first look at the action $L:SU_2\curvearrowright\mathbb{R}^4$ by left multiplication, which is by definition given by:

$$L_U(M)=UM$$

In order to compute the matrix of this action, let us write, as usual:

$$U=xc_1+yc_2+zc_3+tc_4$$

$$M=ac_1+bc_2+cc_3+dc_4$$

By using the multiplication formulae in Theorem 9.11, we obtain:

$$\begin{aligned}UM&=(xc_1+yc_2+zc_3+tc_4)(ac_1+bc_2+cc_3+dc_4)\\&=(xa-yb-zc-td)c_1\\&+(xb+ya+zd-tc)c_2\\&+(xc-yd+za+tb)c_3\\&+(xd+yc-zb+ta)c_4\end{aligned}$$

We conclude that the matrix of the left action considered above is:

$$L_U=\begin{pmatrix}x&-y&-z&-t\\y&x&-t&z\\z&t&x&-y\\t&-z&y&x\end{pmatrix}$$

Similarly, let us look now at the action $R:SU_2\curvearrowright\mathbb{R}^4$ by right multiplication, which is by definition given by the following formula:

$$R_U(M)=MU^*$$

In order to compute the matrix of this action, let us write, as before:

$$U=xc_1+yc_2+zc_3+tc_4$$

$$M=ac_1+bc_2+cc_3+dc_4$$

By using the multiplication formulae in Theorem 9.11, we obtain:

$$\begin{aligned} MU^* &= (ac_1+bc_2+cc_3+dc_4)(xc_1-yc_2-zc_3-tc_4)\\ &= (ax+by+cz+dt)c_1\\ &+ (-ay+bx-ct+dz)c_2\\ &+ (-az+bt+cx-dy)c_3\\ &+ (-at-bz+cy+dx)c_4 \end{aligned}$$

We conclude that the matrix of the right action considered above is:

$$R_U=\begin{pmatrix} x&y&z&t\\ -y&x&-t&z\\ -z&t&x&-y\\ -t&-z&y&x \end{pmatrix}$$

Now by composing, the matrix of the adjoint matrix in the statement is:

$$\begin{aligned} T_U &= R_UL_U\\ &= \begin{pmatrix} x&y&z&t\\ -y&x&-t&z\\ -z&t&x&-y\\ -t&-z&y&x \end{pmatrix}\begin{pmatrix} x&-y&-z&-t\\ y&x&-t&z\\ z&t&x&-y\\ t&-z&y&x \end{pmatrix}\\ &= \begin{pmatrix} 1&0&0&0\\ 0&x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\ 0&2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\ 0&2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2 \end{pmatrix} \end{aligned}$$

Thus, we have indeed the formula in the statement. As for the remaining assertions, these are all clear either from this formula, or from Theorem 9.13. □

We can now formulate the Euler-Rodrigues result, as follows:

Theorem 9.15. *We have a double cover map, obtained via the adjoint representation,*

$$SU_2\to SO_3$$

and this map produces the Euler-Rodrigues formula

$$U=\begin{pmatrix} x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\ 2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\ 2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2 \end{pmatrix}$$

for the generic elements of SO_3.

Proof. We know from the above that we have a group morphism $SU_2\to SO_3$, given by the formula in the statement, and the problem now is that of proving that this is a double cover map, in the sense that it is surjective, and with kernel $\{\pm1\}$.

(1) Regarding the kernel, this is elementary to compute, as follows:

$$\begin{aligned}\ker(SU_2\to SO_3) &= \left\{U\in SU_2\Big|T_U(M)=M,\forall M\in E\right\}\\ &= \left\{U\in SU_2\Big|UM=MU,\forall M\in E\right\}\\ &= \left\{U\in SU_2\Big|Uc_i=c_iU,\forall i\right\}\\ &= \{\pm1\}\end{aligned}$$

(2) Thus, we are done with this, and as a side remark here, this result shows that our morphism $SU_2\to SO_3$ is ultimately a morphism as follows:

$$PU_2\subset SO_3\quad,\quad PU_2=SU_2/\{\pm1\}$$

Here P stands for "projective", and it is possible to say more about the construction $G\to PG$, which can be performed for any subgroup $G\subset U_N$. But we will not get here into this, our next goal being anyway that of proving that we have $PU_2=SO_3$.

(3) We must prove now that the morphism $SU_2\to SO_3$ is surjective. This is something non-trivial, and there are several advanced proofs for this, as follows:

– A first proof is by using Lie theory. To be more precise, the tangent spaces at 1 of both SU_2 and SO_3 can be explicitly computed, by doing some linear algebra, and the morphism $SU_2\to SO_3$ follows to be surjective around 1, and then globally.

– Another proof is via representation theory, as developed following Peter-Weyl. Indeed, the representations of SU_2 and SO_3 are subject to very similar formulae, called Clebsch-Gordan rules, and this shows that $SU_2\to SO_3$ is surjective.

– Yet another advanced proof, which is actually quite bordeline for what can be called "proof", is by using the ADE/McKay classification of the subgroups $G\subset SO_3$, which shows that there is no room strictly inside SO_3 for something as big as PU_2.

(4) In short, with some good knowledge of group theory, we are done. However, this is not our case, and we will present in what follows a more pedestrian proof, which was actually the original proof, based on the fact that any rotation $U\in SO_3$ has an axis.

(5) As a first computation, let us prove that any rotation $U\in Im(SU_2\to SO_3)$ has an axis. We must look for fixed points of such rotations, and by linearity it is enough to look for fixed points belonging to the sphere $S^2_\mathbb{R}\subset\mathbb{R}^3$. Now recall that in our picture for the quotient map $SU_2\to SO_3$, the space $\mathbb{R}^3$ appears as $F=span_\mathbb{R}(c_2,c_3,c_4)$, naturally embedded into the space $\mathbb{R}^4$ appearing as $E=span_\mathbb{R}(c_1,c_2,c_3,c_4)$. Thus, we must look for fixed points belonging to the sphere $S^3_\mathbb{R}\subset\mathbb{R}^4$ whose first coordinate vanishes. But, in our $\mathbb{R}^4=E$ picture, this sphere $S^3_\mathbb{R}$ is the group SU_2. Thus, we must look for fixed points $V\in SU_2$ whose first coordinate with respect to c_1,c_2,c_3,c_4 vanishes, which amounts in saying that the diagonal entries of V must be purely imaginary numbers.

(6) Long story short, via our various identifications, we are led into solving the equation $UV=VU$ with $U,V\in SU_2$, and with V having a purely imaginary diagonal. So, with standard notations for SU_2, we must solve the following equation, with $p\in i\mathbb R$:

$$\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}\begin{pmatrix}p&q\\-\bar{q}&\bar{p}\end{pmatrix}=\begin{pmatrix}p&q\\-\bar{q}&\bar{p}\end{pmatrix}\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}$$

(7) But this is something which is routine. Indeed, by identifying coefficients we obtain the following equations, each appearing twice:

$$b\bar{q}=\bar{b}q\quad,\quad b(p-\bar{p})=(a-\bar{a})q$$

In the case $b=0$ the only equation which is left is $q=0$, and reminding that we must have $p\in i\mathbb R$, we do have solutions, namely two of them, as follows:

$$V=\pm\begin{pmatrix}i&0\\0&i\end{pmatrix}$$

(8) In the remaining case $b\neq 0$, the first equation reads $b\bar{q}\in\mathbb R$, so we must have $q=\lambda b$ with $\lambda\in\mathbb R$. Now with this substitution made, the second equation reads $p-\bar{p}=\lambda(a-\bar{a})$, and since we must have $p\in i\mathbb R$, this gives $2p=\lambda(a-\bar{a})$. Thus, our equations are:

$$q=\lambda b\quad,\quad p=\lambda\cdot\frac{a-\bar{a}}{2}$$

Getting back now to our problem about finding fixed points, assuming $|a|^2+|b|^2=1$ we must find $\lambda\in\mathbb R$ such that the above numbers p,q satisfy $|p|^2+|q|^2=1$. But:

$$\begin{aligned}|p|^2+|q|^2&=|\lambda b|^2+\left|\lambda\cdot\frac{a-\bar{a}}{2}\right|^2\\&=\lambda^2(|b|^2+Im(a)^2)\\&=\lambda^2(1-Re(a)^2)\end{aligned}$$

Thus, we have again two solutions to our fixed point problem, given by:

$$\lambda=\pm\frac{1}{\sqrt{1-Re(a)^2}}$$

(9) Summarizing, we have proved that any rotation $U\in Im(SU_2\to SO_3)$ has an axis, and with the direction of this axis, corresponding to a pair of opposite points on the sphere $S^2_{\mathbb R}\subset\mathbb R^3$, being given by the above formulae, via $S^2_{\mathbb R}\subset S^3_{\mathbb R}=SU_2$.

(10) In order to finish, we must argue that any rotation $U\in SO_3$ has an axis. But this follows for instance from some topology, by using the induced map $S^2_{\mathbb R}\to S^2_{\mathbb R}$. Now since $U\in SO_3$ is uniquely determined by its rotation axis, which can be regarded as a point of $S^2_{\mathbb R}/\{\pm1\}$, plus its rotation angle $t\in[0,2\pi)$, by using $S^2_{\mathbb R}\subset S^3_{\mathbb R}=SU_2$ as in (9) we are led to the conclusion that U is uniquely determined by an element of $SU_2/\{\pm1\}$, and so appears indeed via the Euler-Rodrigues formula, as desired. □

So long for the Euler-Rodrigues formula. As already mentioned, all the above is just the tip of the iceberg, and there are many more things that can be said, which are all interesting, and worth learning. We will be back to this.

Regarding now O_3, the extension from SO_3 is very simple, as follows:

THEOREM 9.16. *We have the Euler-Rodrigues formula*

$$U=\pm\begin{pmatrix}x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2\end{pmatrix}$$

for the generic elements of O_3.

PROOF. This follows from Theorem 9.15, because the determinant of an orthogonal matrix $U\in O_3$ must satisfy $\det U=\pm1$, and in the case $\det U=-1$, we have:

$$\det(-U)=(-1)^3\det U=-\det U=1$$

Thus, assuming $\det U=-1$, we can therefore rescale U into an element $-U\in SO_3$, and this leads to the conclusion in the statement. □

9d. Higher dimensions

Back to arbitrary dimensions, in the real case, we have the following result:

PROPOSITION 9.17. *We have a decomposition as follows, with SO_N^{-1} consisting by definition of the orthogonal matrices having determinant -1:*

$$O_N=SO_N\cup SO_N^{-1}$$

Moreover, when N is odd the set SO_N^{-1} is simply given by $SO_N^{-1}=-SO_N$.

PROOF. The first assertion is clear from definitions, because the determinant of an orthogonal matrix must be ±1. The second assertion is clear too. Finally, when N is even the situation is a bit more complicated, and requires complex numbers. □

In the complex case now, the result is simpler, as follows:

PROPOSITION 9.18. *We have a decomposition as follows, with SU_N^d consisting by definition of the unitary matrices having determinant $d\in\mathbb{T}$:*

$$O_N=\bigcup_{d\in\mathbb{T}}SU_N^d$$

Moreover, the components are $SU_N^d=f\cdot SU_N$, where $f\in\mathbb{T}$ is such that $f^N=d$.

PROOF. This is clear from definitions, and from the fact that the determinant of a unitary matrix belongs to $\mathbb{T}$, by extracting a suitable square root of the determinant. □

It is possible to use the decomposition in Proposition 9.18 in order to say more about what happens in the real case, in the context of Proposition 9.17, but we will not get into this. We will basically stop here with our study of O_N, U_N, and of their versions SO_N, SU_N. As a last result on the subject, however, let us record:

Theorem 9.19. *We have subgroups of O_N, U_N constructed via the condition*

$$(\det U)^d = 1$$

with $d \in \mathbb{N} \cup \{\infty\}$, which generalize both O_N, U_N and SO_N, SU_N.

Proof. This is indeed from definitions, and from the multiplicativity property of the determinant. We will be back to these groups, which are quite specialized, later on. □

Finally, a word about complexification. The passage $O_N \to U_N$ cannot be understood directly, and we must pass here through the corresponding Lie algebras, as follows:

Theorem 9.20. *The passage $O_N \to U_N$ appears via a Lie algebra complexification,*

$$O_N \to \mathfrak{o}_N \to \mathfrak{u}_n \to U_N$$

with the Lie algebra $\mathfrak{u}_N$ being a complexification of the Lie algebra $\mathfrak{o}_N$.

Proof. This is something rather philosophical, the idea being as follows:

(1) The orthogonal and unitary groups O_N, N_N are both Lie groups, and the corresponding Lie algebras $\mathfrak{o}_N, \mathfrak{u}_N$ can be computed by differentiating the equations defining O_N, U_N, with the conclusion being as follows:

$$\mathfrak{o}_N = \left\{ A \in M_N(\mathbb{R}) \Big| A^t = -A \right\}$$

$$\mathfrak{u}_N = \left\{ B \in M_N(\mathbb{C}) \Big| B^* = -B \right\}$$

(2) This was for the correspondences $O_N \to \mathfrak{o}_N$ and $U_N \to \mathfrak{u}_N$. In the other sense, the correspondences $\mathfrak{o}_N \to O_N$ and $\mathfrak{u}_N \to U_N$ appear by exponentiation, the result here stating that, around 1, the orthogonal matrices can be written as $U = e^A$, with $A \in \mathfrak{o}_N$, and the unitary matrices can be written as $U = e^B$, with $B \in \mathfrak{u}_N$.

(3) In view of all this, in order to understand the passage $O_N \to U_N$ it is enough to understand the passage $\mathfrak{o}_N \to \mathfrak{u}_N$. But, in view of the above explicit formulae for $\mathfrak{o}_N, \mathfrak{u}_N$, this is basically an elementary linear algebra problem. Indeed, let us pick an arbitrary matrix $B \in M_N(\mathbb{C})$, and write it as follows, with $A, C \in M_N(\mathbb{R})$:

$$B = A + iC$$

In terms of A, C, the equation $B^* = -B$ defining the Lie algebra $\mathfrak{u}_N$ reads:

$$A^t = -A \quad , \quad C^t = C$$

(4) As a first observation, we must have $A \in \mathfrak{o}_N$. Regarding now C, let us decompose it as follows, with D being its diagonal, and C' being the remainder:

$$C = D + C'$$

The remainder C' being symmetric with 0 on the diagonal, by switching all the signs below the main diagonal we obtain a certain matrix $C'_- \in \mathfrak{o}_N$. Thus, we have decomposed $B \in \mathfrak{u}_N$ as follows, with $A, C' \in \mathfrak{o}_N$, and with $D \in M_N(\mathbb{R})$ being diagonal:

$$B = A + iD + iC'_-$$

(5) As a conclusion now, we have shown that we have a direct sum decomposition of real linear spaces as follows, with $\Delta \subset M_N(\mathbb{R})$ being the diagonal matrices:

$$\mathfrak{u}_N \simeq \mathfrak{o}_N \oplus \Delta \oplus \mathfrak{o}_N$$

Thus, we can stop our study here, and say that we have reached the conclusion in the statement, namely that $\mathfrak{u}_N$ appears as a "complexification" of $\mathfrak{o}_N$. □

9e. Exercises

Exercises:

Exercise 9.21.

Exercise 9.22.

Exercise 9.23.

Exercise 9.24.

Exercise 9.25.

Exercise 9.26.

Exercise 9.27.

Exercise 9.28.

Bonus exercise.

CHAPTER 10

Projective symmetries

10a. Product operations

We discuss in this chapter more complicated symmetry groups. As a starting point here, we have the following result regarding the dihedral group D_N:

THEOREM 10.1. *The dihedral group D_N is the group having $2N$ elements, $R_1,\ldots,R_N$ and $S_1,\ldots,S_N$, called rotations and symmetries, which multiply as follows,*

$$R_kR_l=R_{k+l}$$
$$R_kS_l=S_{k+l}$$
$$S_kR_l=S_{k-l}$$
$$S_kS_l=R_{k-l}$$

with all the indices being taken modulo N.

PROOF. This is something which is self-explanatory, with $R_1,\ldots,R_N$ standing for the rotations of the N-gon, and with $S_1,\ldots,S_N$ standing for the symmetries. □

Observe now that D_N has the same cardinality as $E_N=\mathbb{Z}_N\times\mathbb{Z}_2$. We obviously don't have $D_N\simeq E_N$, because D_N is not abelian, while E_N is. So, our next goal will be that of proving that D_N appears by "twisting" E_N. In order to do this, let us start with:

PROPOSITION 10.2. *The group $E_N=\mathbb{Z}_N\times\mathbb{Z}_2$ is the group having $2N$ elements, $r_1,\ldots,r_N$ and $s_1,\ldots,s_N$, which multiply according to the following rules,*

$$r_kr_l=r_{k+l}$$
$$r_ks_l=s_{k+l}$$
$$s_kr_l=s_{k+l}$$
$$s_ks_l=r_{k+l}$$

with all the indices being taken modulo N.

PROOF. With the notation $\mathbb{Z}_2=\{1,\tau\}$, the elements of the product group $E_N=\mathbb{Z}_N\times\mathbb{Z}_2$ can be labeled $r_1,\ldots,r_N$ and $s_1,\ldots,s_N$, as follows:

$$r_k=(k,1)\quad,\quad s_k=(k,\tau)$$

These elements multiply then according to the formulae in the statement. Now since a group is uniquely determined by its multiplication rules, this gives the result. □

Let us compare now Theorem 10.1 and Proposition 10.2. In order to formally obtain D_N from E_N, we must twist some of the multiplication rules of E_N, namely:

$$s_kr_l = s_{k+l} \to s_{k-l}$$

$$s_ks_l = r_{k+l} \to r_{k-l}$$

Informally, this amounts in following the rule "τ switches the sign of what comes afterwards", and we are led in this way to the following definition:

Definition 10.3. *Given two groups A, G, with an action $A \curvearrowright G$, the crossed product*

$$P = G \rtimes A$$

is the set $G \times A$, with multiplication $(g,a)(h,b) = (gh^a, ab)$.

It is routine to check that P is indeed a group. Observe that when the action is trivial, $h^a = h$ for any $a \in A$ and $h \in H$, we obtain the usual product $G \times A$.

Now with this technology in hand, by getting back to the dihedral group D_N, we can improve Theorem 10.1, into a final result on the subject, as follows:

Theorem 10.4. *We have a crossed product decomposition as follows,*

$$D_N = \mathbb{Z}_N \rtimes \mathbb{Z}_2$$

with $\mathbb{Z}_2 = \{1, \tau\}$ acting on $\mathbb{Z}_N$ via switching signs, $k^\tau = -k$.

Proof. We have an action $\mathbb{Z}_2 \curvearrowright \mathbb{Z}_N$ given by the formula in the statement, namely $k^\tau = -k$, so we can consider the corresponding crossed product group:

$$P_N = \mathbb{Z}_N \rtimes \mathbb{Z}_2$$

In order to understand the structure of P_N, we follow Proposition 10.2. The elements of P_N can indeed be labeled $\rho_1, \ldots, \rho_N$ and $\sigma_1, \ldots, \sigma_N$, as follows:

$$\rho_k = (k, 1) \quad , \quad \sigma_k = (k, \tau)$$

Now when computing the products of such elements, we basically obtain the formulae in Proposition 10.2, perturbed as in Definition 10.3. To be more precise, we have:

$$\rho_k\rho_l = \rho_{k+l}$$

$$\rho_k\sigma_l = \sigma_{k+l}$$

$$\sigma_k\rho_l = \sigma_{k+l}$$

$$\sigma_k\sigma_l = \rho_{k+l}$$

But these are exactly the multiplication formulae for D_N, from Theorem 10.1. Thus, we have an isomorphism $D_N \simeq P_N$ given by $R_k \to \rho_k$ and $S_k \to \sigma_k$, as desired. □

More generally now, for the transitive graphs, that we are mostly interested in, the point is that at very small values of the order, $N = 2, \ldots, 9$, these always decompose as products, via three main types of graph products, constructed as follows:

Definition 10.5. *Given two finite graphs X,Y, we can construct:*

(1) *The direct product $X\times Y$ has vertex set $X\times Y$, and edges:*

$$(i,\alpha)-(j,\beta)\iff i-j,\ \alpha-\beta$$

(2) *The Cartesian product $X\,\square\,Y$ has vertex set $X\times Y$, and edges:*

$$(i,\alpha)-(j,\beta)\iff i=j,\ \alpha-\beta\ \text{or}\ i-j,\alpha=\beta$$

(3) *The lexicographic product $X\circ Y$ has vertex set $X\times Y$, and edges:*

$$(i,\alpha)-(j,\beta)\iff \alpha-\beta\ \text{or}\ \alpha=\beta,\ i-j$$

We call these three products the standard products of graphs.

Several comments can be made here. First, the direct product $X\times Y$ is the usual one in a categorical sense, and we will leave clarifying this observation as an exercise. The Cartesian product $X\,\square\,Y$ is quite natural too from a geometric perspective, for instance because a product by a segment gives a prism. As for the lexicographic product $X\circ Y$, this is something interesting too, obtained by putting a copy of X at each vertex of Y.

At the level of symmetry groups, several things can be said, and we first have:

Theorem 10.6. *We have group embeddings as follows, for any graphs X,Y,*

$$G(X)\times G(Y)\subset G(X\times Y)$$

$$G(X)\times G(Y)\subset G(X\,\square\,Y)$$

$$G(X)\wr G(Y)\subset G(X\circ Y)$$

but these embeddings are not always isomorphisms.

Proof. The fact that we have indeed embeddings as above is clear from definitions. As for the counterexamples, in each case, these are easy to construct as well, provided by our study of small graphs, at $N=2,\ldots,11$, and we will leave this as an exercise. □

The problem now is that of deciding when the embeddings in Theorem 10.6 are isomorphisms. In order to discusss this, we first have the following basic fact:

Theorem 10.7. *Given a subgroup $G\subset S_N$, regarded as matrix group via*

$$G\subset S_N\subset O_N$$

the standard coordinates of the group elements, $u_{ij}(g)=g_{ij}$, are given by:

$$u_{ij}=\chi\left(\sigma\in G\Big|\sigma(j)=i\right)$$

Moreover, these functions $u_{ij}:G\to\mathbb{C}$ generate the algebra $C(G)$.

Proof. Here the first assertion comes from the fact that the entries of the permutation matrices $\sigma \in S_N \subset O_N$, acting as $\sigma(e_i) = e_{\sigma(i)}$, are given by the following formula:

$$\sigma_{ij} = \begin{cases} 1 & \text{if } \sigma(j) = i \\ 0 & \text{otherwise} \end{cases}$$

As for the second assertion, this comes from the Stone-Weierstrass theorem, because the coordinate functions $u_{ij} : G \to \mathbb{C}$ obviously separate the group elements $\sigma \in G$. □

We are led in this way to the following definition:

Definition 10.8. *The magic matrix associated to a permutation group $G \subset S_N$ is the $N \times N$ matrix of characteristic functions*

$$u_{ij} = \chi\left(\sigma \in G\middle|\sigma(j) = i\right)$$

with the name "magic" coming from the fact that, on each row and each column, these characteristic functions sum up to 1.

The interest in this notion comes from the fact, that we know from Theorem 10.7, that the entries of the magic matrix generate the algebra of functions on our group:

$$C(G) = < u_{ij} >$$

We will talk more in detail later about such matrices, and their correspondence with the subgroups $G \subset S_N$, and what can be done with it, in the general framework of representation theory. However, for making our point, here is the general principle:

Principle 10.9. *Everything that you can do with your group $G \subset S_N$ can be expressed in terms of the magic matrix $u = (u_{ij})$, quite often with good results.*

This principle comes from the above Stone-Weierstrass result, $C(G) = < u_{ij} >$. Indeed, when coupled with some basic spectral theory, and more specifically with the Gelfand theorem from operator algebras, this result tells us that our group G appears as the spectrum of the algebra $< u_{ij} >$, therefore leading to the above principle.

As an illustration for all this, in relation with the graphs, we have:

Theorem 10.10. *Given a subgroup $G \subset S_N$, the transpose of its action map $X \times G \to X$ on the set $X = \{1, \ldots, N\}$, given by $(i, \sigma) \to \sigma(i)$, is given by:*

$$\Phi(e_i) = \sum_j e_j \otimes u_{ji}$$

Also, in the case where we have a graph with N vertices, the action of G on the vertex set X leaves invariant the edges precisely when we have

$$du = ud$$

with d being as usual the adjacency matrix of the graph.

Proof. There are several things going on here, the idea being as follows:

(1) Given a subgroup $G\subset S_N$, if we set $X=\{1,\ldots,N\}$, we have indeed an action map as follows, and with the reasons of using $X\times G$ instead of the perhaps more familiar $G\times X$ being dictated by some quantum algebra, that we will do later in this book:

$$a:X\times G\to X\quad,\quad a(i,\sigma)=\sigma(i)$$

(2) Now by transposing this map, we obtain a morphism of algebras, as follows:

$$\Phi:C(X)\to C(X)\otimes C(G)\quad,\quad \Phi(f)(i,\sigma)=f(\sigma(i))$$

When evaluated on the Dirac masses, this map Φ is then given by:

$$\Phi(e_i)(j,\sigma)=e_i(\sigma(j))=\delta_{\sigma(j)i}$$

Thus, in tensor product notation, we have the following formula, as desired:

$$\Phi(e_i)(j,\sigma)=\left(\sum_j e_j\otimes u_{ji}\right)(j,\sigma)$$

(3) Regarding now the second assertion, observe first that we have:

$$(du)_{ij}(\sigma)=\sum_k d_{ik}u_{kj}(\sigma)=\sum_k d_{ik}\delta_{\sigma(j)k}=d_{i\sigma(j)}$$

On the other hand, we have as well the following formula:

$$(ud)_{ij}(\sigma)=\sum_k u_{ik}(\sigma)d_{kj}=\sum_k \delta_{\sigma(k)i}d_{kj}=d_{\sigma^{-1}(i)j}$$

Thus $du=ud$ reformulates as $d_{ij}=d_{\sigma(i)\sigma(j)}$, which gives the result. □

Back to graphs, we want to know when the embeddings in Theorem 10.6 are isomorphisms. In what regards the first two products, we have here the following result:

Theorem 10.11. *Let X and Y be finite connected regular graphs. If their spectra $\{\lambda\}$ and $\{\mu\}$ do not contain 0 and satisfy*

$$\left\{\lambda_i/\lambda_j\right\}\cap\left\{\mu_k/\mu_l\right\}=\{1\}$$

then $G(X\times Y)=G(X)\times G(Y)$. Also, if their spectra satisfy

$$\left\{\lambda_i-\lambda_j\right\}\cap\left\{\mu_k-\mu_l\right\}=\{0\}$$

then $G(X\,\square\,Y)=G(X)\times G(Y)$.

Proof. This is something quite standard, the idea being as follows:

(1) First, we know from Theorem 10.6 that we have embeddings as follows, valid for any two graphs X,Y, and coming from definitions:

$$G(X)\times G(Y)\subset G(X\times Y)$$
$$G(X)\times G(Y)\subset G(X\,\square\,Y)$$

(2) Now let λ_1 be the valence of X. Since X is regular we have $\lambda_1 \in Sp(X)$, with 1 as eigenvector, and since X is connected λ_1 has multiplicity 1. Thus if P_1 is the orthogonal projection onto $\mathbb{C}1$, the spectral decomposition of d_X is of the following form:

$$d_X = \lambda_1 P_1 + \sum_{i \neq 1} \lambda_i P_i$$

We have a similar formula for the adjacency matrix d_Y, namely:

$$d_Y = \mu_1 Q_1 + \sum_{j \neq 1} \mu_j Q_j$$

(3) But this gives the following formulae for the graph products:

$$d_{X \times Y} = \sum_{ij} (\lambda_i \mu_j) P_i \otimes Q_j$$

$$d_{X \,\square\, Y} = \sum_{ij} (\lambda_i + \mu_i) P_i \otimes Q_j$$

Here the projections form partitions of unity, and the scalar are distinct, so these are spectral decompositions. The coactions will commute with any of the spectral projections, and so with both $P_1 \otimes 1$, $1 \otimes Q_1$. In both cases the universal coaction v is the tensor product of its restrictions to the images of $P_1 \otimes 1$, $1 \otimes Q_1$, which gives the result. □

Regarding now the lexicographic product, things here are more tricky. Let us first recall that the lexicographic product of two graphs $X \circ Y$ is obtained by putting a copy of X at each vertex of Y, the formula for the edges being as follows:

$$(i, \alpha) - (j, \beta) \Longleftrightarrow \alpha - \beta \text{ or } \alpha = \beta,\ i - j$$

In what regards now the computation of the symmetry group, as before we must do here some spectral theory, and we are led in this way to the following result:

Theorem 10.12. *Let X, Y be regular graphs, with X connected. If their spectra $\{\lambda_i\}$ and $\{\mu_j\}$ satisfy the condition*

$$\left\{\lambda_1 - \lambda_i \middle| i \neq 1\right\} \cap \left\{ - n\mu_j\right\} = \emptyset$$

where n and λ_1 are the order and valence of X, then $G(X \circ Y) = G(X) \wr G(Y)$.

Proof. This is something quite tricky, the idea being as follows:

(1) First, we know from Theorem 10.6 that we have an embedding as follows, valid for any two graphs X, Y, and coming from definitions:

$$G(X) \wr G(Y) \subset G(X \circ Y)$$

(2) We denote by P_i, Q_j the spectral projections corresponding to λ_i, μ_j. Since X is connected we have $P_1 = \mathbb{I}/n$, and we obtain:

$$\begin{aligned} d_{X \circ Y} &= d_X \otimes 1 + \mathbb{I} \otimes d_Y \\ &= \left(\sum_i \lambda_i P_i\right) \otimes \left(\sum_j Q_j\right) + (nP_1) \otimes \left(\sum_i \mu_j Q_j\right) \\ &= \sum_j (\lambda_1 + n\mu_j)(P_1 \otimes Q_j) + \sum_{i \neq 1} \lambda_i (P_i \otimes 1) \end{aligned}$$

In this formula the projections form a partition of unity and the scalars are distinct, so this is the spectral decomposition of $d_{X \circ Y}$.

(3) Now let W be the universal magic matrix for $X \circ Y$. Then W must commute with all spectral projections, and in particular:

$$[W, P_1 \otimes Q_j] = 0$$

Summing over j gives $[W, P_1 \otimes 1] = 0$, so $1 \otimes C(Y)$ is invariant under the coaction. So, consider the restriction of W, which gives a coaction of $G(X \circ Y)$ on $1 \otimes C(Y)$, that we can denote as follows, with y being a certain magic unitary:

$$W(1 \otimes e_a) = \sum_b 1 \otimes e_b \otimes y_{ba}$$

(4) On the other hand, according to our definition of W, we can write:

$$W(e_i \otimes 1) = \sum_{jb} e_j \otimes e_b \otimes x_{ji}^b$$

By multiplying by the previous relation, found in (3), we obtain:

$$\begin{aligned} W(e_i \otimes e_a) &= \sum_{jb} e_j \otimes e_b \otimes y_{ba} x_{ji}^b \\ &= \sum_{jb} e_j \otimes e_b \otimes x_{ji}^b y_{ba} \end{aligned}$$

But this shows that the coefficients of W are of the following form:

$$W_{jb,ia} = y_{ba} x_{ji}^b = x_{ji}^b y_{ba}$$

(5) In order to advance, consider now the following matrix:

$$x^b = (x_{ij}^b)$$

Since the map W above is a morphism of algebras, each row of x^b is a partition of unity. Also, by using the antipode map S, which is transpose to $g\to g^{-1}$, we have:

$$\begin{aligned}S\left(\sum_j x^b_{ji}\right)&=S\left(\sum_{ja}x^b_{ji}y_{ba}\right)\\&=S\left(\sum_{ja}W_{jb,ia}\right)\\&=\sum_{ja}W_{ia,jb}\\&=\sum_{ja}x^a_{ij}y_{ab}\\&=\sum_a y_{ab}\\&=1\end{aligned}$$

(6) We check now that both x^a,y commute with d_X,d_Y. We have:

$$(d_{X\circ Y})_{ia,jb}=(d_X)_{ij}\delta_{ab}+(d_Y)_{ab}$$

Thus the two products between W and $d_{X\circ Y}$ are given by:

$$(Wd_{X\circ Y})_{ia,kc}=\sum_j W_{ia,jc}(d_X)_{jk}+\sum_{jb}W_{ia,jb}(d_Y)_{bc}$$

$$(d_{X\circ Y}W)_{ia,kc}=\sum_j (d_X)_{ij}W_{ja,kc}+\sum_{jb}(d_Y)_{ab}W_{jb,kc}$$

(7) Now since the magic matrix W commutes by definition with $d_{X\circ Y}$, the terms on the right in the above equations are equal, and by summing over c we get:

$$\sum_j x^a_{ij}(d_X)_{jk}+\sum_{cb}y_{ab}(d_Y)_{bc}=\sum_j (d_X)_{ij}x^a_{jk}+\sum_{cb}(d_Y)_{ab}y_{bc}$$

The second sums in both terms are equal to the valence of Y, so we get $[x^a,d_X]=0$. Now once again from the formula coming from $[W,d_{X\circ Y}]=0$, we get:

$$[y,d_Y]=0$$

(8) Summing up, the coefficients of W are of the following form, where x^b are magic unitaries commuting with d_X, and y is a magic unitary commuting with d_Y:

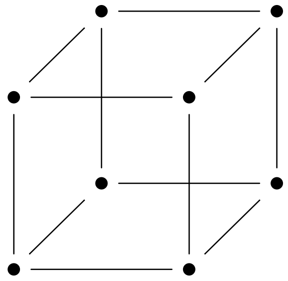
$$W_{jb,ia}=x^b_{ji}y_{ba}$$

But this gives a morphism $C(G(X)\wr G(Y))\to G(X\circ Y)$ mapping $u^{(b)}_{ji}\to x^b_{ji}$ and $v_{ba}\to y_{ba}$, which is inverse to the morphism in (1), as desired. □

10b. Hyperoctahedral groups

At a more advanced level now, we first have the hyperoctahedral group H_N. This group is something quite tricky, which appears as follows:

Definition 10.13. *The hyperoctahedral group H_N is the group of symmetries of the unit cube in $\mathbb{R}^N$,*



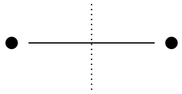
viewed as a graph, or equivalently, as a metric space.

Here the equivalence at the end is clear from definitions, because any symmetry of the cube graph must preserve the lengths of the edges, and so we have:

$$G(\square_{graph}) = G(\square_{metric})$$

The hyperoctahedral group is a quite interesting group, whose definition, as a symmetry group, reminds that of the dihedral group D_N. So, let us start our study in the same way as we did for D_N, with a discussion at small values of $N \in \mathbb{N}$:

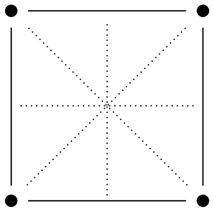
$\underline{N=1}$. Here the 1-cube is the segment, whose symmetries are the identity id, plus the symmetry τ with respect to the middle of the segment:



Thus, we obtain the group with 2 elements, which is a very familiar object:

$$H_1 = D_2 = S_2 = \mathbb{Z}_2$$

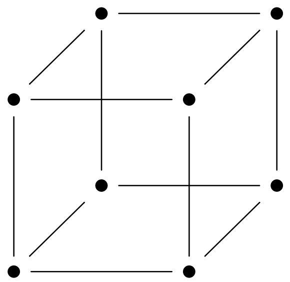
$\underline{N=2}$. Here the 2-cube is the square, whose symmetries are the 4 rotations, of angles $0°, 90°, 180°, 270°$, and the 4 symmetries with respect to the 4 symmetry axes, which are the 2 diagonals, and the 2 segments joining the midpoints of opposite sides:



Thus, we obtain a group with 8 elements, which again is a very familiar object:

$$H_2 = D_4 = \mathbb{Z}_4 \rtimes \mathbb{Z}_2$$

$\underline{N=3}$. Here the 3-cube is the usual cube in $\mathbb{R}^3$, pictured as follows:



However, in relation with the symmetries, the situation now is considerably more complicated, because, thinking well, this cube has no less than 48 symmetries. Precisely identifying and counting these symmetries is actually an excellent exercise.

All this looks quite complicated, but fortunately we can count H_N, at $N = 3$, and at higher N as well, by using some tricks, the result being as follows:

THEOREM 10.14. *We have the cardinality formula*

$$|H_N| = 2^N N!$$

coming from the fact that H_N is the symmetry group of the coordinate axes of $\mathbb{R}^N$.

PROOF. This follows from some geometric thinking, as follows:

(1) Consider the standard cube in $\mathbb{R}^N$, centered at 0, and having as vertices the points having coordinates ± 1. With this picture in hand, it is clear that the symmetries of the cube coincide with the symmetries of the N coordinate axes of $\mathbb{R}^N$.

(2) In order to count now these latter symmetries, a bit as we did for the dihedral group, observe first that we have $N!$ permutations of these N coordinate axes.

(3) But each of these permutations of the coordinate axes $\sigma \in S_N$ can be further "decorated" by a sign vector $e \in \{\pm 1\}^N$, consisting of the possible ± 1 flips which can be applied to each coordinate axis, at the arrival.

(4) And the point is that, obviously, we obtain in this way all the elements of H_N. Thus, we have the following formula, for the cardinality of H_N:

$$|H_N| = |S_N| \cdot |\mathbb{Z}_2^N| = N! \cdot 2^N$$

Thus, we are led to the conclusions in the statement. □

As in the dihedral group case, it is possible to go beyond this, with a crossed product decomposition, of quite special type, called wreath product decomposition:

Theorem 10.15. *We have a wreath product decomposition as follows,*

$$H_N=\mathbb{Z}_2\wr S_N$$

which means by definition that we have a crossed product decomposition

$$H_N=\mathbb{Z}_2^N\rtimes S_N$$

with the permutations $\sigma\in S_N$ acting on the elements $e\in\mathbb{Z}_2^N$ as follows:

$$\sigma(e_1,\ldots,e_k)=(e_{\sigma(1)},\ldots,e_{\sigma(k)})$$

In particular we have, as found before, the cardinality formula $|H_N|=2^NN!$.

Proof. As explained in the proof of Theorem 10.14, the elements of H_N can be identified with the pairs $g=(e,\sigma)$ consisting of a permutation $\sigma\in S_N$, and a sign vector $e\in\mathbb{Z}_2^N$, so that at the level of the cardinalities, we have the following formula:

$$|H_N|=|\mathbb{Z}_2^N\times S_N|$$

To be more precise, given an element $g\in H_N$, the element $\sigma\in S_N$ is the corresponding permutation of the N coordinate axes, regarded as unoriented lines in $\mathbb{R}^N$, and $e\in\mathbb{Z}_2^N$ is the vector collecting the possible flips of these coordinate axes, at the arrival. Now observe that the product formula for two such pairs $g=(e,\sigma)$ is as follows, with the permutations $\sigma\in S_N$ acting on the elements $f\in\mathbb{Z}_2^N$ as in the statement:

$$(e,\sigma)(f,\tau)=(ef^\sigma,\sigma\tau)$$

Thus, we are precisely in the framework of the crossed products, as constructed in chapter 1, and we conclude that we have a crossed product decomposition, as follows:

$$H_N=\mathbb{Z}_2^N\rtimes S_N$$

Thus, we are led to the conclusion in the statement, with the formula $H_N=\mathbb{Z}_2\wr S_N$ being just a shorthand for the decomposition $H_N=\mathbb{Z}_2^N\rtimes S_N$ that we found. □

We will be back to the hyperoctahedral groups later on, on several occasions, with further results about them, both of algebraic and of analytic type.

10c. Complex reflections

The groups that we studied so far are all groups of orthogonal matrices. When looking into general unitary matrices, we led to the following interesting class of groups:

Definition 10.16. *The complex reflection group $H_N^s\subset U_N$, depending on parameters*

$$N\in\mathbb{N}\quad,\quad s\in\mathbb{N}\cup\{\infty\}$$

is the group of permutation-type matrices with s-th roots of unity as entries,

$$H_N^s=M_N(\mathbb{Z}_s\cup\{0\})\cap U_N$$

with the convention $\mathbb{Z}_\infty=\mathbb{T}$, at $s=\infty$.

This construction is something quite tricky, that will keep as busy, for the remainder of this section. As a first observation, at $s=1,2$ we obtain the following groups:

$$H_N^1=S_N\quad,\quad H_N^2=H_N$$

Another important particular case of the above construction is $s=\infty$, where we obtain a group which is actually not finite, but is still compact, denoted as follows:

$$K_N\subset U_N$$

This latter group K_N is called full complex reflection group, and will appear many times, in what follows. Let us summarize now these observations, as follows:

Proposition 10.17. *The complex reflection groups $H_N^s\subset U_N$ are as follows:*

1. *At $s=1$ we have $H_N^1=S_N$, having cardinality $|S_N|=N!$.*
2. *At $s=2$ we have $H_N^2=H_N$, having cardinality $|H_N|=2^NN!$.*
3. *At $s=\infty$ we have $H_N^\infty=K_N$, having cardinality $|K_N|=\infty$.*

Proof. This is clear indeed from the above discussion, and with the cardinality results at $s=1$ and $s=2$ being something that we know well. □

Let us record as well the following result, which is elementary too:

Proposition 10.18. *We have inclusions as follows, for any r,s:*

$$r|s\implies H_r\subset H_s$$

In particular, we have inclusions $S_N\subset H_N^s\subset K_N$, for any s.

Proof. With the cyclic group $\mathbb{Z}_s$ being viewed as group of the s-th roots of unity, in the complex plane, as in Definition 10.16, we have inclusions as follows:

$$r|s\implies \mathbb{Z}_r\subset\mathbb{Z}_s$$

Thu, with the group H_N^s constructed as in Definition 10.16, for $r|s$ we have:

$$\begin{aligned}H_N^r&=M_N(\mathbb{Z}_r\cup\{0\})\cap U_N\\&\subset M_N(\mathbb{Z}_s\cup\{0\})\cap U_N\\&=H_N^s\end{aligned}$$

Finally, the last assertion is clear, and comes as well from this, since for any s:

$$1|s|\infty$$

Thus, we are led to the conclusions in the statement. □

In general, in analogy with what we know about S_N,H_N, we first have:

Proposition 10.19. *The number of elements of H_N^s with $s\in\mathbb{N}$ is:*

$$|H_N^s|=s^NN!$$

At $s=\infty$, the group $K_N=H_N^\infty$ that we obtain is infinite.

Proof. This is indeed clear from our definition of H_N^s, as a matrix group as above, because there are $N!$ choices for a permutation-type matrix, and then s^N choices for the corresponding s-roots of unity, which must decorate the N nonzero entries. □

Once again in analogy with what we know at $s = 1, 2$, we have as well:

Theorem 10.20. *We have a wreath product decomposition*

$$H_N^s = \mathbb{Z}_s^N \rtimes S_N = \mathbb{Z}_s \wr S_N$$

with the permutations $\sigma \in S_N$ acting on the elements $e \in \mathbb{Z}_s^N$ as follows:

$$\sigma(e_1, \ldots, e_k) = (e_{\sigma(1)}, \ldots, e_{\sigma(k)})$$

In particular we have, as found before, the cardinality formula $|H_N^s| = s^N N!$.

Proof. As explained in the proof of Proposition 10.19, the elements of H_N^s can be identified with the pairs $g = (e, \sigma)$ consisting of a permutation $\sigma \in S_N$, and a decorating vector $e \in \mathbb{Z}_s^N$, so that at the level of the cardinalities, we have:

$$|H_N| = |\mathbb{Z}_s^N \times S_N|$$

Now observe that the product formula for two such pairs $g = (e, \sigma)$ is as follows, with the permutations $\sigma \in S_N$ acting on the elements $f \in \mathbb{Z}_s^N$ as in the statement:

$$(e, \sigma)(f, \tau) = (ef^\sigma, \sigma\tau)$$

Thus, we are in the framework of the crossed products, and we obtain $H_N^s = \mathbb{Z}_s^N \rtimes S_N$. But this can be written, by definition, as $H_N^s = \mathbb{Z}_s \wr S_N$, and we are done. □

Finally, in relation with graph symmetries, the above groups appear as follows:

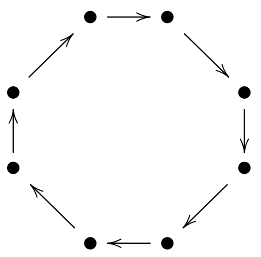
Theorem 10.21. *The complex reflection group H_N^s appears as symmetry group,*

$$H_N^s = G(NC_s)$$

with NC_s consisting of N disjoint copies of the oriented cycle C_s.

Proof. This is something elementary, the idea being as follows:

(1) Consider first the oriented cycle C_s, which looks as follows:



It is then clear that the symmetry group of this graph is the cyclic group $\mathbb{Z}_s$.

(2) In the general case now, where we have $N\in\mathbb N$ disjoint copies of the above cycle C_s, we must suitably combine the corresponding N copies of the cyclic group $\mathbb Z_s$. But this leads to the wreath product group $H_N^s=\mathbb Z_s\wr S_N$, as stated. □

10d. Reflection groups

Back to the rotation groups, in the real case, we have the following result:

Proposition 10.22. *We have a decomposition as follows, with SO_N^{-1} consisting by definition of the orthogonal matrices having determinant -1:*

$$O_N=SO_N\cup SO_N^{-1}$$

Moreover, when N is odd the set SO_N^{-1} is simply given by $SO_N^{-1}=-SO_N$.

Proof. The first assertion is clear from definitions, because the determinant of an orthogonal matrix must be ±1. The second assertion is clear too. Finally, when N is even the situation is a bit more complicated, and requires complex numbers. □

In the complex case now, the result is simpler, as follows:

Proposition 10.23. *We have a decomposition as follows, with SU_N^d consisting by definition of the unitary matrices having determinant $d\in\mathbb T$:*

$$O_N=\bigcup_{d\in\mathbb T}SU_N^d$$

Moreover, the components are $SU_N^d=f\cdot SU_N$, where $f\in\mathbb T$ is such that $f^N=d$.

Proof. This is clear from definitions, and from the fact that the determinant of a unitary matrix belongs to $\mathbb T$, by extracting a suitable square root of the determinant. □

It is possible to use the decomposition in Proposition 10.23 in order to say more about what happens in the real case, in the context of Proposition 10.22, but we will not get into this. We will basically stop here with our study of O_N,U_N, and of their versions SO_N,SU_N. As a last result on the subject, however, let us record:

Theorem 10.24. *We have subgroups of O_N,U_N constructed via the condition*

$$(\det U)^d=1$$

with $d\in\mathbb N\cup\{\infty\}$, which generalize both O_N,U_N and SO_N,SU_N.

Proof. This is indeed from definitions, and from the multiplicativity property of the determinant. We will be back to these groups, which are quite specialized, later on. □

With this discussed, let us go back now to the complex reflection groups from the previous section, and make a link with the material there. We first have:

Theorem 10.25. *The full complex reflection group $K_N\subset U_N$, given by*

$$K_N=M_N(\mathbb T\cup\{0\})\cap U_N$$

has a wreath product decomposition as follows,

$$K_N=\mathbb T\wr S_N$$

with S_N acting on $\mathbb T^N$ in the standard way, by permuting the factors.

Proof. This is something that we know from before, appearing as the $s=\infty$ particular case of the results established there for the complex reflection groups H_N^s. □

By using the above full complex reflection group K_N, we can talk in fact about the reflection subgroup of any compact group $G\subset U_N$, as follows:

Definition 10.26. *Given $G\subset U_N$, we define its reflection subgroup to be*

$$K=G\cap K_N$$

with the intersection taken inside U_N.

This notion is something quite interesting, leading us into the question of understanding what the subgroups of K_N are. We have here the following construction:

Theorem 10.27. *We have subgroups of the basic complex reflection groups,*

$$H_N^{sd}\subset H_N^s$$

constructed via the following condition, with $d\in\mathbb N\cup\{\infty\}$,

$$(\det U)^d=1$$

which generalize all the complex reflection groups that we have so far.

Proof. Here the first assertion is clear from definitions, and from the multiplicativity of the determinant. As for the second assertion, this is rather a remark, coming from the fact that the alternating group A_N, which is the only finite group so far not fitting into the series $\{H_N^s\}$, is indeed of this type, obtained from $H_N^1=S_N$ by using $d=1$. □

The point now is that, by a well-known and deep result in group theory, the complex reflection groups consist of the series $\{H_N^{sd}\}$ constructed above, and of a number of exceptional groups, which can be fully classified. To be more precise, we have:

Theorem 10.28. *The irreducible complex reflection groups are*

$$H_N^{sd}=\left\{U\in H_N^s\Big|(\det U)^d=1\right\}$$

along with 34 exceptional examples.

Proof. This is something quite advanced, and we refer here to the paper of Shephard and Todd, and to the subsequent literature on the subject. □

10e. Exercises

Exercises:

Exercise 10.29.

Exercise 10.30.

Exercise 10.31.

Exercise 10.32.

Exercise 10.33.

Exercise 10.34.

Exercise 10.35.

Exercise 10.36.

Bonus exercise.

CHAPTER 11

Representation theory

11a. Representations

Time now for some more advanced mathematics. Following Weyl, we have:

Definition 11.1. *A unitary representation of a compact group G is a continuous group morphism into a unitary group*

$$v:G\to U_N\quad,\quad g\to v_g$$

which can be faithful or not. The character of such a representation is the function

$$\chi:G\to\mathbb C\quad,\quad g\to Tr(v_g)$$

where Tr is the usual, unnormalized trace of the $N\times N$ matrices.

At the level of examples, most of the compact groups that we met so far, finite or continuous, naturally appear as closed subgroups $G\subset U_N$. In this case, the embedding $G\subset U_N$ is of course a representation, called fundamental representation. In general now, let us first discuss the various operations on the representations. We have here:

Proposition 11.2. *The representations of a compact group G are subject to:*

1. *Making sums. Given representations v,w, of dimensions N,M, their sum is the $N+M$-dimensional representation $v+w=diag(v,w)$.*
2. *Making products. Given representations v,w, of dimensions N,M, their product is the NM-dimensional representation $(v\otimes w)_{ia,jb}=v_{ij}w_{ab}$.*
3. *Taking conjugates. Given a N-dimensional representation v, its conjugate is the N-dimensional representation $(\bar v)_{ij}=\bar v_{ij}$.*
4. *Spinning by unitaries. Given a N-dimensional representation v, and a unitary $U\in U_N$, we can spin v by this unitary, $v\to UvU^*$.*

Proof. The fact that the operations in the statement are indeed well-defined, among morphisms from G to unitary groups, is indeed clear from definitions. □

In relation now with characters, we have the following result:

Proposition 11.3. *We have the following formulae, regarding characters*

$$\chi_{v+w}=\chi_v+\chi_w\quad,\quad\chi_{v\otimes w}=\chi_v\chi_w\quad,\quad\chi_{\bar v}=\bar\chi_v\quad,\quad\chi_{UvU^*}=\chi_v$$

in relation with the basic operations for the representations.

Proof. All these assertions are elementary, by using the following well-known trace formulae, valid for any square matrices V,W, and any unitary U:

$$Tr(diag(V,W))=Tr(V)+Tr(W)\quad,\quad Tr(V\otimes W)=Tr(V)Tr(W)$$

$$Tr(\bar{V})=\overline{Tr(V)}\quad,\quad Tr(UVU^*)=Tr(V)$$

Thus, we are led to the formulae in the statement. □

Assume now that we are given a closed subgroup $G\subset U_N$. By using the above operations, we can construct a whole family of representations of G, as follows:

Definition 11.4. *Given a closed subgroup $G\subset U_N$, its Peter-Weyl representations are the various tensor products between the fundamental representation and its conjugate:*

$$v:G\subset U_N\quad,\quad \bar{v}:G\subset U_N$$

We denote these tensor products $v^{\otimes k}$, with $k=\circ\bullet\bullet\circ\ldots$ being a colored integer, with the colored tensor powers being defined according to the rules

$$v^{\otimes\circ}=v\quad,\quad v^{\otimes\bullet}=\bar{v}\quad,\quad v^{\otimes kl}=v^{\otimes k}\otimes v^{\otimes l}$$

and with the convention that $v^{\otimes\emptyset}$ is the trivial representation $1:G\to U_1$.

Here are a few examples of such representations, namely those coming from the colored integers of length 2, which will often appear in what follows:

$$v^{\otimes\circ\circ}=v\otimes v\quad,\quad v^{\otimes\circ\bullet}=v\otimes\bar{v}$$

$$v^{\otimes\bullet\circ}=\bar{v}\otimes v\quad,\quad v^{\otimes\bullet\bullet}=\bar{v}\otimes\bar{v}$$

In relation now with characters, we have the following result:

Proposition 11.5. *The characters of the Peter-Weyl representations are given by*

$$\chi_{v^{\otimes k}}=(\chi_v)^k$$

with the colored powers being given by $\chi^\circ=\chi$, $\chi^\bullet=\bar{\chi}$ and multiplicativity.

Proof. This follows indeed from the additivity, multiplicativity and conjugation formulae from Proposition 11.3, via the conventions in Definition 11.4. □

Getting back now to our motivations, we can see the interest in the above constructions. Indeed, the joint moments of the main character $\chi=\chi_v$ and its adjoint $\bar{\chi}=\chi_{\bar{v}}$ are the expectations of the characters of various Peter-Weyl representations:

$$\int_G\chi^k=\int_G\chi_{v^{\otimes k}}$$

In order to advance, we must develop some general theory. Let us start with:

Definition 11.6. *Given a compact group G, and two of its representations,*

$$v:G\to U_N\quad,\quad w:G\to U_M$$

we define the space of intertwiners between these representations as being

$$Hom(v,w)=\left\{T\in M_{M\times N}(\mathbb C)\Big|Tv_g=w_gT,\forall g\in G\right\}$$

and we use the following conventions:

1. *We use the notations $Fix(v)=Hom(1,v)$, and $End(v)=Hom(v,v)$.*
2. *We write $v\sim w$ when $Hom(v,w)$ contains an invertible element.*
3. *We say that v is irreducible, and write $v\in Irr(G)$, when $End(v)=\mathbb C1$.*

Here the terminology is something very standard, with Fix, Hom, End standing respectively for the fixed points, homomorphisms and endomorphisms. We will see later that irreducible means indecomposable, in a suitable sense.

Here are now a few basic results, regarding the above spaces:

Theorem 11.7. *The spaces of intertwiners have the following properties:*

1. *$T\in Hom(v,w),S\in Hom(w,z)\implies ST\in Hom(v,z)$.*
2. *$S\in Hom(v,w),T\in Hom(z,t)\implies S\otimes T\in Hom(v\otimes z,w\otimes t)$.*
3. *$T\in Hom(v,w)\implies T^*\in Hom(w,v)$.*

In abstract terms, we say that the Hom spaces form a tensor $*$-category.

Proof. All the formulae in the statement are indeed clear from definitions, via elementary computations. As for the last assertion, this is something coming from (1,2,3). We will be back to tensor categories later on, with more details on this latter fact. □

As a main consequence of the above result, we have:

Theorem 11.8. *Given a representation $v:G\to U_N$, the linear space*

$$End(v)\subset M_N(\mathbb C)$$

is a $*$-algebra, with respect to the usual involution of the matrices.

Proof. By definition, $End(v)$ is a linear subspace of $M_N(\mathbb C)$. We know from Proposition 11.7 (1) that this subspace $End(v)$ is a subalgebra of $M_N(\mathbb C)$, and then we know as well from Proposition 11.7 (3) that this subalgebra is stable under the involution $*$. Thus, what we have here is a $*$-subalgebra of $M_N(\mathbb C)$, as claimed. □

11b. Peter-Weyl

In order to exploit Theorem 11.8, we will need a basic result from linear algebra, stating that any $*$-algebra $A \subset M_N(\mathbb{C})$ decomposes as a direct sum, as follows:

$$A \simeq M_{N_1}(\mathbb{C}) \oplus \ldots \oplus M_{N_k}(\mathbb{C})$$

Indeed, let us write the unit $1 \in A$ as $1 = p_1 + \ldots + p_k$, with $p_i \in A$ being central minimal projections. Then each of the spaces $A_i = p_i A p_i$ is a subalgebra of A, and we have a decomposition $A = A_1 \oplus \ldots \oplus A_k$. But since each central projection $p_i \in A$ was chosen minimal, we have $A_i \simeq M_{N_i}(\mathbb{C})$, with $N_i = rank(p_i)$, as desired.

We can now formulate our first Peter-Weyl type theorem, as follows:

Theorem 11.9 (Peter-Weyl 1). *Let $v : G \to U_N$ be a representation, consider the algebra $A = End(v)$, and write its unit $1 = p_1 + \ldots + p_k$ as above. We have then*

$$v = v_1 + \ldots + v_k$$

with each v_i being an irreducible representation, obtained by restricting v to $Im(p_i)$.

Proof. This basically follows from Theorem 11.8, as follows:

(1) We first associate to our representation $v : G \to U_N$ the corresponding action map on $\mathbb{C}^N$. If a linear subspace $W \subset \mathbb{C}^N$ is invariant, the restriction of the action map to W is an action map too, which must come from a subrepresentation $w \subset v$.

(2) Consider now a projection $p \in End(v)$. From $pv = vp$ we obtain that the linear space $W = Im(p)$ is invariant under v, and so this space must come from a subrepresentation $w \subset v$. It is routine to check that the operation $p \to w$ maps subprojections to subrepresentations, and minimal projections to irreducible representations.

(3) With these preliminaries in hand, let us decompose the algebra $End(v)$ as above, by using the decomposition $1 = p_1 + \ldots + p_k$ into central minimal projections. If we denote by $v_i \subset v$ the subrepresentation coming from the vector space $V_i = Im(p_i)$, then we obtain in this way a decomposition $v = v_1 + \ldots + v_k$, as in the statement. □

Here is now our second Peter-Weyl theorem, complementing Theorem 11.9:

Theorem 11.10 (Peter-Weyl 2). *Given a closed subgroup $G \subset_v U_N$, any of its irreducible smooth representations*

$$w : G \to U_M$$

appears inside a tensor product of the fundamental representation v and its adjoint $\bar{v}$.

Proof. Given a representation $w : G \to U_M$, we define the space of coefficients $C_w \subset C(G)$ of this representation as being the following linear space:

$$C_w = span\Big[g \to w(g)_{ij}\Big]$$

With this notion in hand, the result can be deduced as follows:

(1) The construction $w \to C_w$ is functorial, in the sense that it maps subrepresentations into linear subspaces. This is indeed something which is routine to check.

(2) A closed subgroup $G \subset_v U_N$ is a Lie group, and a representation $w : G \to U_M$ is smooth when we have an inclusion $C_w \subset < C_v >$. This is indeed well-known.

(3) By definition of the Peter-Weyl representations, as arbitrary tensor products between the fundamental representation v and its conjugate $\bar{v}$, we have:

$$< C_v >= \sum_k C_{v^{\otimes k}}$$

(4) Now by putting together the above observations (2,3) we conclude that we must have an inclusion as follows, for certain exponents $k_1, \ldots, k_p$:

$$C_w \subset C_{v^{\otimes k_1} \oplus \ldots \oplus v^{\otimes k_p}}$$

(5) By using now (1), we deduce that we have an inclusion $w \subset v^{\otimes k_1} \oplus \ldots \oplus v^{\otimes k_p}$, and by applying Theorem 14.10, this leads to the conclusion in the statement. □

In order to further advance with Peter-Weyl theory, we need to talk about integration over G. In the finite group case the situation is trivial, as follows:

Proposition 11.11. *Any finite group G has a unique probability measure which is invariant under left and right translations,*

$$\mu(E) = \mu(gE) = \mu(Eg)$$

and this is the normalized counting measure on G, given by $\mu(E) = |E|/|G|$.

Proof. This is indeed something trivial, which follows from definitions. □

In the general, continuous case, let us begin with the following key result:

Proposition 11.12. *Given a unital positive linear form $\psi : C(G) \to \mathbb{C}$, the limit*

$$\int_\varphi f = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \psi^{*k}(f)$$

exists, and for a coefficient of a representation $f = (\tau \otimes id)w$ we have

$$\int_\varphi f = \tau(P)$$

where P is the orthogonal projection onto the 1-eigenspace of $(id \otimes \psi)w$.

Proof. By linearity it is enough to prove the first assertion for functions of the following type, where w is a Peter-Weyl representation, and τ is a linear form:

$$f=(\tau\otimes id)w$$

Thus we are led into the second assertion, and more precisely we can have the whole result proved if we can establish the following formula, with $f=(\tau\otimes id)w$:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\psi^{*k}(f)=\tau(P)$$

In order to prove this latter formula, observe that we have:

$$\psi^{*k}(f)=(\tau\otimes\psi^{*k})w=\tau((id\otimes\psi^{*k})w)$$

Let us set $M=(id\otimes\psi)w$. In terms of this matrix, we have:

$$((id\otimes\psi^{*k})w)_{i_0i_{k+1}}=\sum_{i_1\ldots i_k}M_{i_0i_1}\ldots M_{i_ki_{k+1}}=(M^k)_{i_0i_{k+1}}$$

Thus we have the following formula, valid for any $k\in\mathbb{N}$:

$$(id\otimes\psi^{*k})w=M^k$$

It follows that our Cesàro limit is given by the following formula:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\psi^{*k}(f)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\tau(M^k)=\tau\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^nM^k\right)$$

Now since w is unitary we have $||w||=1$, and so $||M||\leq1$. Thus the last Cesàro limit converges, and equals the orthogonal projection onto the 1-eigenspace of M:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^nM^k=P$$

Thus our initial Cesàro limit converges as well, to $\tau(P)$, as desired. □

When the linear form $\psi\in C(G)^*$ is faithful, we have the following finer result:

Proposition 11.13. *Given a faithful unital linear form $\psi\in C(G)^*$, the limit*

$$\int_\psi f=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\psi^{*k}(f)$$

exists, and is independent of ψ, given on coefficients of representations by

$$\left(id\otimes\int_\psi\right)w=P$$

where P is the orthogonal projection onto the space $Fix(w)=\left\{\xi\in\mathbb{C}^n\middle|w\xi=\xi\right\}$.

PROOF. In view of Proposition 11.12, it remains to prove that when ψ is faithful, the 1-eigenspace of the matrix $M = (id \otimes \psi)w$ equals the space $Fix(w)$.

"$\supset$" This is clear, and for any ψ, because we have the following implication:

$$w\xi = \xi \implies M\xi = \xi$$

"$\subset$" Here we must prove that, when ψ is faithful, we have:

$$M\xi = \xi \implies w\xi = \xi$$

For this purpose, assume that we have $M\xi = \xi$, and consider the following function:

$$f = \sum_i \left(\sum_j w_{ij}\xi_j - \xi_i\right)\left(\sum_k w_{ik}\xi_k - \xi_i\right)^*$$

We must prove that we have $f = 0$. Since v is unitary, we have:

$$\begin{aligned} f &= \sum_{ijk} w_{ij}w_{ik}^*\xi_j\bar{\xi}_k - \frac{1}{N}w_{ij}\xi_j\bar{\xi}_i - \frac{1}{N}w_{ik}^*\xi_i\bar{\xi}_k + \frac{1}{N^2}\xi_i\bar{\xi}_i \\ &= \sum_j |\xi_j|^2 - \sum_{ij} w_{ij}\xi_j\bar{\xi}_i - \sum_{ik} w_{ik}^*\xi_i\bar{\xi}_k + \sum_i |\xi_i|^2 \\ &= ||\xi||^2 - < w\xi, \xi > - \overline{< w\xi, \xi >} + ||\xi||^2 \\ &= 2(||\xi||^2 - Re(< w\xi, \xi >)) \end{aligned}$$

By using now our assumption $M\xi = \xi$, we obtain from this:

$$\begin{aligned} \psi(f) &= 2\psi(||\xi||^2 - Re(< w\xi, \xi >)) \\ &= 2(||\xi||^2 - Re(< M\xi, \xi >)) \\ &= 2(||\xi||^2 - ||\xi||^2) \\ &= 0 \end{aligned}$$

Now since ψ is faithful, this gives $f = 0$, and so $w\xi = \xi$, as claimed. □

We can now formulate a main result, as follows:

THEOREM 11.14. *Any compact group G has a unique Haar integration, which can be constructed by starting with any faithful positive unital form $\psi \in C(G)^*$, and setting:*

$$\int_G = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \psi^{*k}$$

Moreover, for any representation w we have the formula

$$\left(id \otimes \int_G\right) w = P$$

where P is the orthogonal projection onto $Fix(w) = \left\{\xi \in \mathbb{C}^n \middle| w\xi = \xi\right\}$.

Proof. Let us first go back to the general context of Proposition 11.12. Since convolving one more time with ψ will not change the Cesàro limit appearing there, the functional $\int_\psi\in C(G)^*$ constructed there has the following invariance property:

$$\int_\psi*\psi=\psi*\int_\psi=\int_\psi$$

In the case where ψ is assumed to be faithful, as in Proposition 11.13, our claim is that we have the following formula, valid this time for any $\varphi\in C(G)^*$:

$$\int_\psi*\varphi=\varphi*\int_\psi=\varphi(1)\int_\psi$$

Indeed, it is enough to prove this formula on a coefficient of a corepresentation:

$$f=(\tau\otimes id)w$$

In order to do so, consider the following two matrices:

$$P=\left(id\otimes\int_\psi\right)w\quad,\quad Q=(id\otimes\varphi)w$$

We have then the following formulae, which all follow from definitions:

$$\left(\int_\psi*\varphi\right)f=\tau(PQ)\quad,\quad\left(\varphi*\int_\psi\right)f=\tau(QP)\quad,\quad\varphi(1)\int_\psi f=\varphi(1)\tau(P)$$

Thus, in order to prove our claim, it is enough to establish the following formula:

$$PQ=QP=\psi(1)P$$

But this follows from the fact, from Proposition 11.13, that $P=(id\otimes\int_\psi)w$ is the orthogonal projection onto $Fix(w)$. Thus, we proved our claim. Now observe that, with $\Delta f(g\otimes h)=f(gh)$, this formula that we proved can be written as follows:

$$\varphi\left(\int_\psi\otimes id\right)\Delta=\varphi\left(id\otimes\int_\psi\right)\Delta=\varphi\int_\psi(.)1$$

This formula being true for any $\varphi\in C(G)^*$, we can simply delete φ, and we conclude that $\int_G=\int_\psi$ has the required left and right invariance property, namely:

$$\left(\int_G\otimes id\right)\Delta=\left(id\otimes\int_G\right)\Delta=\int_G(.)1$$

Finally, the uniqueness is clear as well, because if we have two invariant integrals $\int_G,\int_G'$, then their convolution equals on one hand $\int_G$, and on the other hand, $\int_G'$. □

Summarizing, we know how to integrate over G. Before getting into probabilistic applications, let us develop however more Peter-Weyl theory. We will need:

Proposition 11.15. *We have a Frobenius type isomorphism*

$$Hom(v,w)\simeq Fix(v\otimes\bar{w})$$

valid for any two representations v,w.

Proof. According to definitions, we have the following equivalences:

$$\begin{aligned}T\in Hom(v,w)&\iff Tv=wT\\&\iff \sum_i T_{ai}v_{ij}=\sum_b w_{ab}T_{bj},\forall a,j\end{aligned}$$

On the other hand, we have as well the following equivalences:

$$\begin{aligned}T\in Fix(v\otimes\bar{w})&\iff (v\otimes\bar{w})T=\xi\\&\iff \sum_{bi}v_{ji}\bar{w}_{ab}T_{bi}=T_{aj}\forall a,j\end{aligned}$$

With these formulae in hand, both inclusions follow from the unitarity of v,w. □

We can now formulate a third Peter-Weyl theorem, as follows:

Theorem 11.16 (Peter-Weyl 3). *The dense subalgebra $\mathcal{C}(G)\subset C(G)$ generated by the coefficients of the fundamental representation decomposes as a direct sum*

$$\mathcal{C}(G)=\bigoplus_{w\in Irr(G)}M_{\dim(w)}(\mathbb{C})$$

with the summands being pairwise orthogonal with respect to the scalar product

$$<f,g>=\int_G f\bar{g}$$

where $\int_G$ is the Haar integration over G.

Proof. By combining the previous two Peter-Weyl results, Theorems 11.9 and 11.10, we deduce that we have a linear space decomposition as follows:

$$\mathcal{C}(G)=\sum_{w\in Irr(G)}C_w=\sum_{w\in Irr(G)}M_{\dim(w)}(\mathbb{C})$$

Thus, in order to conclude, it is enough to prove that for any two irreducible representations $v,w\in Irr(G)$, the corresponding spaces of coefficients are orthogonal:

$$v\not\sim w\implies C_v\perp C_w$$

But this follows from Theorem 11.14, via Proposition 11.15. Let us set indeed:

$$P_{ia,jb}=\int_G v_{ij}\bar{w}_{ab}$$

Then P is the orthogonal projection onto the following vector space:

$$Fix(v \otimes \bar{w}) \simeq Hom(v, w) = \{0\}$$

Thus we have $P = 0$, and this gives the result. □

Finally, we have the following result, completing the Peter-Weyl theory:

Theorem 11.17 (Peter-Weyl 4). *The characters of irreducible representations belong to the algebra*

$$\mathcal{C}(G)_{central} = \left\{ f \in \mathcal{C}(G) \Big| f(gh) = f(hg), \forall g, h \in G \right\}$$

called algebra of central functions on G, and form an orthonormal basis of it.

Proof. Observe first that $\mathcal{C}(G)_{central}$ is indeed an algebra, which contains all the characters. Conversely, consider a function $f \in \mathcal{C}(G)$, written as follows:

$$f = \sum_{w \in Irr(G)} f_w$$

The condition $f \in \mathcal{C}(G)_{central}$ states then that for any $w \in Irr(G)$, we must have:

$$f_w \in \mathcal{C}(G)_{central}$$

But this means that f_w must be a scalar multiple of χ_w, so the characters form a basis of $\mathcal{C}(G)_{central}$, as stated. Also, the fact that we have an orthogonal basis follows from Theorem 11.16. As for the fact that the characters have norm 1, this follows from:

$$\int_G \chi_w \bar{\chi}_w = \sum_{ij} \int_G w_{ii} \bar{w}_{jj} = \sum_i \frac{1}{M} = 1$$

Here we have used the fact, coming from Theorem 11.14 and Proposition 11.15, that the integrals $\int_G w_{ij} \bar{w}_{kl}$ form the orthogonal projection onto the following vector space:

$$Fix(w \otimes \bar{w}) \simeq End(w) = \mathbb{C}1$$

Thus, the proof of our theorem is now complete. □

As a key observation here, complementing Theorem 11.17, observe that a function $f : G \to \mathbb{C}$ is central, in the sense that it satisfies $f(gh) = f(hg)$, precisely when it satisfies the following condition, saying that it must be constant on conjugacy classes:

$$f(ghg^{-1}) = f(h), \forall g, h \in G$$

Thus, in the finite group case for instance, the algebra of central functions is something which is very easy to compute, and this gives useful information about $Rep(G)$. We will not get into this here, but some of our exercises will be about this.

As a basic illustration now for all this, we have the following result:

Theorem 11.18. *For a compact abelian group G the irreducible representations are all 1-dimensional, and form the dual discrete abelian group $\widehat{G}$.*

Proof. This is clear from the Peter-Weyl theory, because when G is abelian any function $f:G\to\mathbb{C}$ is central, and so the algebra of central functions is $\mathcal{C}(G)$ itself, and so the irreducible representations $u\in Irr(G)$ coincide with their characters $\chi_u\in\widehat{G}$. □

Many other things can be said, along these lines.

11c. Clebsch-Gordan

As a last piece of Lie group theory, we are now in position of dealing, in a quite conceptual way, with SU_2 and SO_3. Regarding SU_2, the result here is as follows:

Theorem 11.19. *The irreducible representations of SU_2 are all self-adjoint, and can be labelled by positive integers, with their fusion rules being as follows,*

$$r_k\otimes r_l=r_{|k-l|}+r_{|k-l|+2}+\ldots+r_{k+l}$$

called Clebsch-Gordan rules. The corresponding dimensions are $\dim r_k=k+1$.

Proof. There are several proofs for this fact, the simplest one, with the knowledge that we have, being via purely algebraic methods, as follows:

(1) Our first claim is that we have the following estimate, telling us that the even moments of the main character are smaller than the Catalan numbers:

$$\int_{SU_2}\chi^{2k}\leq C_k$$

But this is something which is elementary, obtained by using $SU_2\simeq S^3_\mathbb{R}$ and standard spherical integrals, and with the stronger statement that we have in fact equality $=$. However, for the purposes of what follows, the above $\leq$ estimate will do.

(2) Alternatively, the above estimate can be deduced with purely algebraic methods, by using an easiness type argument for SU_2, as follows:

$$\begin{aligned}\int_{SU_2}\chi^{2k}&=\dim(Fix(u^{\otimes 2k}))\\&=\dim\left(span\left(T'_\pi\Big|\pi\in NC_2(2k)\right)\right)\\&\leq|NC_2(2k)|\\&=C_k\end{aligned}$$

To be more precise, SU_2 is not exactly easy, but rather "super-easy", coming from a different implementation $\pi\to T'_\pi$ of the pairings, involving some signs. And with this being proved exactly as the Brauer theorem for O_N, with modifications where needed.

(3) Long story short, we have our estimate in (1), and this is all that we need. Our claim is that we can construct, by recurrence on $k \in \mathbb{N}$, a sequence r_k of irreducible, self-adjoint and distinct representations of SU_2, satisfying:

$$r_0 = 1 \quad , \quad r_1 = u \quad , \quad r_k + r_{k-2} = r_{k-1} \otimes r_1$$

Indeed, assume that $r_0, \ldots, r_{k-1}$ are constructed, and let us construct r_k. We have:

$$r_{k-1} + r_{k-3} = r_{k-2} \otimes r_1$$

Thus $r_{k-1} \subset r_{k-2} \otimes r_1$, and since r_{k-2} is irreducible, by Frobenius we have:

$$r_{k-2} \subset r_{k-1} \otimes r_1$$

We conclude there exists a certain representation r_k such that:

$$r_k + r_{k-2} = r_{k-1} \otimes r_1$$

(4) By recurrence, r_k is self-adjoint. Now observe that according to our recurrence formula, we can split $u^{\otimes k}$ as a sum of the following type, with positive coefficients:

$$u^{\otimes k} = c_k r_k + c_{k-2} r_{k-2} + \ldots$$

We conclude by Peter-Weyl that we have an inequality as follows, with equality precisely when r_k is irreducible, and non-equivalent to the other summands r_i:

$$\sum_i c_i^2 \leq \dim(End(u^{\otimes k}))$$

(5) But by (1) the number on the right is $\leq C_k$, and some straightforward combinatorics, based on the fusion rules, shows that the number on the left is C_k as well:

$$C_k = \sum_i c_i^2 \leq \dim(End(u^{\otimes k})) = \int_{SU_2} \chi^{2k} \leq C_k$$

Thus we have equality in our estimate, so our representation r_k is irreducible, and non-equivalent to $r_{k-2}, r_{k-4}, \ldots$ Moreover, this representation r_k is not equivalent to $r_{k-1}, r_{k-3}, \ldots$ either, with this coming from $r_p \subset u^{\otimes p}$ for any p, and from:

$$\dim(Fix(u^{\otimes 2s+1})) = \int_{SU_2} \chi^{2s+1} = 0$$

(6) Thus, we proved our claim. Now since each irreducible representation of SU_2 appears into some $u^{\otimes k}$, and we know how to decompose each $u^{\otimes k}$ into sums of representations r_k, these representations r_k are all the irreducible representations of SU_2, and we are done with the main assertion. As for the dimension formula, this is clear. □

Regarding now SO_3, we have here a similar result, as follows:

Theorem 11.20. *The irreducible representations of SO_3 are all self-adjoint, and can be labelled by positive integers, with their fusion rules being as follows,*

$$r_k \otimes r_l = r_{|k-l|} + r_{|k-l|+1} + \ldots + r_{k+l}$$

also called Clebsch-Gordan rules. The corresponding dimensions are $\dim r_k = 2k + 1$.

Proof. As before with SU_2, there are many possible proofs here, which are all instructive. Here is our take on the subject, in the spirit of our proof for SU_2:

(1) Our first claim is that we have the following formula, telling us that the moments of the main character equal the Catalan numbers:

$$\int_{SO_3} \chi^k = C_k$$

But this is something that we know from before, coming from Euler-Rodrigues. Alternatively, this can be deduced as well from Tannakian duality, a bit as for SU_2.

(2) Our claim now is that we can construct, by recurrence on $k \in \mathbb{N}$, a sequence r_k of irreducible, self-adjoint and distinct representations of SO_3, satisfying:

$$r_0 = 1 \quad , \quad r_1 = u - 1 \quad , \quad r_k + r_{k-1} + r_{k-2} = r_{k-1} \otimes r_1$$

Indeed, assume that $r_0, \ldots, r_{k-1}$ are constructed, and let us construct r_k. The Frobenius trick from the proof for SU_2 will no longer work, due to some technical reasons, so we have to invoke (1). To be more precise, by integrating characters we obtain:

$$r_{k-1}, r_{k-2} \subset r_{k-1} \otimes r_1$$

Thus there exists a representation r_k such that:

$$r_{k-1} \otimes r_1 = r_k + r_{k-1} + r_{k-2}$$

(3) Once again by integrating characters, we conclude that r_k is irreducible, and non-equivalent to $r_1, \ldots, r_{k-1}$, and this proves our claim. Also, since any irreducible representation of SO_3 must appear in some tensor power of u, and we can decompose each $u^{\otimes k}$ into sums of representations r_p, we conclude that these representations r_p are all the irreducible representations of SO_3. Finally, the dimension formula is clear. □

There are of course many other things that can be said about SU_2 and SO_3. For instance, with the proof of Theorem 11.19 and Theorem 11.20 done in a purely algebraic fashion, by using the super-easiness property of SU_2 and SO_3, the Euler-Rodrigues formula can be deduced afterwards from this, without any single computation, the argument being that by Peter-Weyl the embedding $PU_2 \subset SO_3$ must be indeed an equality.

11d. McKay subgroups

McKay subgroups.

11e. Exercises

Exercises:

Exercise 11.21.

Exercise 11.22.

Exercise 11.23.

Exercise 11.24.

Exercise 11.25.

Exercise 11.26.

Exercise 11.27.

Exercise 11.28.

Bonus exercise.

CHAPTER 12

Easiness, diagrams

12a. Easy groups

Let us formulate the following key definition, extending to the case of arbitrary partitions what we already know from chapter 11 about pairings:

Definition 12.1. *Given a partition $\pi\in P(k,l)$ and an integer $N\in\mathbb N$, we define*

$$T_\pi:(\mathbb C^N)^{\otimes k}\to(\mathbb C^N)^{\otimes l}$$

by the following formula, with $e_1,\ldots,e_N$ being the standard basis of $\mathbb C^N$,

$$T_\pi(e_{i_1}\otimes\ldots\otimes e_{i_k})=\sum_{j_1\ldots j_l}\delta_\pi\begin{pmatrix}i_1&\ldots&i_k\\ j_1&\ldots&j_l\end{pmatrix}e_{j_1}\otimes\ldots\otimes e_{j_l}$$

and with the coefficients on the right being Kronecker type symbols.

To be more precise here, in order to compute the Kronecker type symbols $\delta_\pi(^i_j)\in\{0,1\}$, we proceed exactly as in the pairing case, namely by putting the multi-indices $i=(i_1,\ldots,i_k)$ and $j=(j_1,\ldots,j_l)$ on the legs of π, in the obvious way. In case all the blocks of π contain equal indices of i,j, we set $\delta_\pi(^i_j)=1$. Otherwise, we set $\delta_\pi(^i_j)=0$.

With the above notion in hand, we can now formulate the following key definition, motivated by the Brauer theorems for O_N,U_N, as indicated before:

Definition 12.2. *A closed subgroup $G\subset U_N$ is called easy when*

$$Hom(u^{\otimes k},u^{\otimes l})=span\left(T_\pi\Big|\pi\in D(k,l)\right)$$

for any two colored integers $k,l=\circ\bullet\circ\bullet\ldots$, for certain sets of partitions

$$D(k,l)\subset P(k,l)$$

where $\pi\to T_\pi$ is the standard implementation of the partitions, as linear maps.

In other words, we call a group G easy when its Tannakian category appears in the simplest possible way: from the linear maps associated to partitions. The terminology is quite natural, because Tannakian duality is basically our only serious tool.

As basic examples, the orthogonal and unitary groups O_N, U_N are both easy, coming respectively from the following collections of sets of partitions:

$$P_2=\bigsqcup_{k,l}P_2(k,l)\quad,\quad\mathcal{P}_2=\bigsqcup_{k,l}\mathcal{P}_2(k,l)$$

In the general case now, as an important theoretical remark, in the context of Definition 12.2, consider the following collection of sets of partitions:

$$D=\bigsqcup_{k,l}D(k,l)$$

This collection of sets D obviously determines G, but the converse is not true. Indeed, at $N=1$ for instance, both the choices $D=P_2,\mathcal{P}_2$ produce the same easy group, namely $G=\{1\}$. We will be back to this issue on several occasions, with results about it.

In order to advance, our first goal will be that of establishing a duality between easy groups and certain special classes of collections of sets as above, namely:

$$D=\bigsqcup_{k,l}D(k,l)$$

Let us begin with a general definition, as follows:

Definition 12.3. *Let $P(k,l)$ be the set of partitions between an upper colored integer k, and a lower colored integer l. A collection of subsets*

$$D=\bigsqcup_{k,l}D(k,l)$$

with $D(k,l)\subset P(k,l)$ is called a category of partitions when it has the following properties:

1. *Stability under the horizontal concatenation, $(\pi,\sigma)\to[\pi\sigma]$.*
2. *Stability under vertical concatenation $(\pi,\sigma)\to[^\sigma_\pi]$, with matching middle symbols.*
3. *Stability under the upside-down turning $*$, with switching of colors, $\circ\leftrightarrow\bullet$.*
4. *Each set $P(k,k)$ contains the identity partition $||\ldots||$.*
5. *The sets $P(\emptyset,\circ\bullet)$ and $P(\emptyset,\bullet\circ)$ both contain the semicircle $\cap$.*
6. *The sets $P(k,\bar{k})$ with $|k|=2$ contain the crossing partition $\slash\!\!\!\backslash$.*

As before, this is something that we already met in chapter 11, but for the pairings only. Observe the similarity with the axioms for Tannakian categories, also from chapter 11. We will see in a moment that this similarity can be turned into something very precise, the idea being that such a category produces a family of easy quantum groups $(G_N)_{N\in\mathbb{N}}$, one for each $N\in\mathbb{N}$, via the formula in Definition 12.1, and Tannakian duality.

As basic examples, that we have already met in chapter 6, in connection with the representation theory of O_N,U_N, we have the categories $P_2,\mathcal{P}_2$ of pairings, and of matching pairings. Further basic examples include the categories P,P_{even} of all partitions, and of all partitions whose blocks have even size. We will see in a moment that these latter categories are related to the symmetric and hyperoctahedral groups S_N, H_N.

The relation with the Tannakian categories comes from the following result:

Proposition 12.4. *The assignement $\pi \to T_\pi$ is categorical, in the sense that*

$$T_\pi \otimes T_\sigma = T_{[\pi\sigma]} \quad , \quad T_\pi T_\sigma = N^{c(\pi,\sigma)} T_{[^\sigma_\pi]} \quad , \quad T_\pi^* = T_{\pi^*}$$

where $c(\pi,\sigma)$ are certain integers, coming from the erased components in the middle.

Proof. The concatenation axiom follows from the following computation:

$$\begin{aligned}
&(T_\pi \otimes T_\sigma)(e_{i_1} \otimes \ldots \otimes e_{i_p} \otimes e_{k_1} \otimes \ldots \otimes e_{k_r})\\
=&\sum_{j_1\ldots j_q}\sum_{l_1\ldots l_s} \delta_\pi\begin{pmatrix}i_1&\ldots&i_p\\ j_1&\ldots&j_q\end{pmatrix}\delta_\sigma\begin{pmatrix}k_1&\ldots&k_r\\ l_1&\ldots&l_s\end{pmatrix} e_{j_1}\otimes\ldots\otimes e_{j_q}\otimes e_{l_1}\otimes\ldots\otimes e_{l_s}\\
=&\sum_{j_1\ldots j_q}\sum_{l_1\ldots l_s} \delta_{[\pi\sigma]}\begin{pmatrix}i_1&\ldots&i_p&k_1&\ldots&k_r\\ j_1&\ldots&j_q&l_1&\ldots&l_s\end{pmatrix} e_{j_1}\otimes\ldots\otimes e_{j_q}\otimes e_{l_1}\otimes\ldots\otimes e_{l_s}\\
=&\ T_{[\pi\sigma]}(e_{i_1} \otimes \ldots \otimes e_{i_p} \otimes e_{k_1} \otimes \ldots \otimes e_{k_r})
\end{aligned}$$

The composition axiom follows from the following computation:

$$\begin{aligned}
&T_\pi T_\sigma(e_{i_1}\otimes\ldots\otimes e_{i_p})\\
=&\sum_{j_1\ldots j_q}\delta_\sigma\begin{pmatrix}i_1&\ldots&i_p\\ j_1&\ldots&j_q\end{pmatrix}\sum_{k_1\ldots k_r}\delta_\pi\begin{pmatrix}j_1&\ldots&j_q\\ k_1&\ldots&k_r\end{pmatrix} e_{k_1}\otimes\ldots\otimes e_{k_r}\\
=&\sum_{k_1\ldots k_r} N^{c(\pi,\sigma)}\delta_{[^\sigma_\pi]}\begin{pmatrix}i_1&\ldots&i_p\\ k_1&\ldots&k_r\end{pmatrix} e_{k_1}\otimes\ldots\otimes e_{k_r}\\
=&\ N^{c(\pi,\sigma)} T_{[^\sigma_\pi]}(e_{i_1}\otimes\ldots\otimes e_{i_p})
\end{aligned}$$

Finally, the involution axiom follows from the following computation:

$$\begin{aligned}
&T_\pi^*(e_{j_1}\otimes\ldots\otimes e_{j_q})\\
=&\sum_{i_1\ldots i_p} < T_\pi^*(e_{j_1}\otimes\ldots\otimes e_{j_q}), e_{i_1}\otimes\ldots\otimes e_{i_p} > e_{i_1}\otimes\ldots\otimes e_{i_p}\\
=&\sum_{i_1\ldots i_p}\delta_\pi\begin{pmatrix}i_1&\ldots&i_p\\ j_1&\ldots&j_q\end{pmatrix} e_{i_1}\otimes\ldots\otimes e_{i_p}\\
=&\ T_{\pi^*}(e_{j_1}\otimes\ldots\otimes e_{j_q})
\end{aligned}$$

Summarizing, our correspondence is indeed categorical. □

Time now to put everyting together. All the above was pure combinatorics, and in relation with the compact groups, we have the following result:

Theorem 12.5. *Each category of partitions $D = (D(k,l))$ produces a family of compact groups $G = (G_N)$, one for each $N \in \mathbb{N}$, via the formula*

$$Hom(u^{\otimes k}, u^{\otimes l}) = span\left(T_\pi \Big| \pi \in D(k,l)\right)$$

and the Tannakian duality correspondence.

Proof. Given an integer $N \in \mathbb{N}$, consider the correspondence $\pi \to T_\pi$ constructed in Definition 12.1, and then the collection of linear spaces in the statement, namely:

$$C_{kl} = span\left(T_\pi \Big| \pi \in D(k,l)\right)$$

According to the formulae in Proposition 12.4, and to our axioms for the categories of partitions, from Definition 12.3, this collection of spaces $C = (C_{kl})$ satisfies the axioms for the Tannakian categories, from chapter 11. Thus the Tannakian duality result there applies, and provides us with a closed subgroup $G_N \subset U_N$ such that:

$$C_{kl} = Hom(u^{\otimes k}, u^{\otimes l})$$

Thus, we are led to the conclusion in the statement. □

In relation with the easiness property, we can now formulate a key result, which can serve as an alternative definition for the easy groups, as follows:

Theorem 12.6. *A closed subgroup $G \subset U_N$ is easy precisely when*

$$Hom(u^{\otimes k}, u^{\otimes l}) = span\left(T_\pi \Big| \pi \in D(k,l)\right)$$

for any colored integers k, l, for a certain category of partitions $D \subset P$.

Proof. This basically follows from Theorem 12.5, as follows:

(1) In one sense, we know from Theorem 12.5 that any category of partitions $D \subset P$ produces a family of closed groups $G \subset U_N$, one for each $N \in \mathbb{N}$, according to Tannakian duality and to the Hom space formula there, namely:

$$Hom(u^{\otimes k}, u^{\otimes l}) = span\left(T_\pi \Big| \pi \in D(k,l)\right)$$

But these groups $G \subset U_N$ are indeed easy, in the sense of Definition 12.2.

(2) In the other sense now, assume that $G \subset U_N$ is easy, in the sense of Definition 12.2, coming via the above Hom space formula, from a collection of sets as follows:

$$D = \bigsqcup_{k,l} D(k,l)$$

Consider now the category of partitions $\widetilde{D} =< D >$ generated by this family. This is by definition the smallest category of partitions containing D, whose existence follows by starting with D, and performing the various categorical operations, namely horizontal and vertical concatenation, and upside-down turning. It follows then, via another application of Tannakian duality, that we have the following formula, for any k,l:

$$Hom(u^{\otimes k},u^{\otimes l})=span\left(T_\pi\Big|\pi\in\widetilde{D}(k,l)\right)$$

Thus, our group $G\subset U_N$ can be viewed as well as coming from $\widetilde{D}$, and so appearing as particular case of the construction in Theorem 12.5, and this gives the result. □

As already mentioned above, Theorem 12.6 can be regarded as an alternative definition for easiness, with the assumption that $D\subset P$ must be a category of partitions being added. In what follows we will rather use this new definition, which is more precise.

Generally speaking, the same comments as before apply. First, G is easy when its Tannakian category appears in the simplest possible way: from a category of partitions. The terminology is quite natural, because Tannakian duality is our only serious tool.

Also, the category of partitions D is not unique, for instance because at $N=1$ all the categories of partitions produce the same easy group, namely $G=\{1\}$. We will be back to this issue on several occasions, with various results about it.

We will see in what follows that many interesting examples of compact quantum groups are easy. Moreover, most of the known series of "basic" compact quantum groups, $G=(G_N)$ with $N\in\mathbb{N}$, can be in principle made fit into some suitable extensions of the easy quantum group formalism. We will discuss this too, in what follows.

The notion of easiness goes back to the results of Brauer regarding the orthogonal group O_N, and the unitary group U_N, which reformulate as follows:

THEOREM 12.7. *We have the following results:*

1. *The unitary group U_N is easy, coming from the category $\mathcal{P}_2$.*
2. *The orthogonal group O_N is easy as well, coming from the category P_2.*

PROOF. This is something that we already know, from chapter 11, based on Tannakian duality, the idea of the proof being as follows:

(1) The group U_N being defined via the relations $u^*=u^{-1}$, $u^t=\bar{u}^{-1}$, the associated Tannakian category is $C=span(T_\pi|\pi\in D)$, with:

$$D=<\overset{\cap}{\circ\bullet}\,,\ \overset{\cap}{\bullet\circ}>=\mathcal{P}_2$$

(2) The group $O_N\subset U_N$ being defined by imposing the relations $u_{ij}=\bar{u}_{ij}$, the associated Tannakian category is $C=span(T_\pi|\pi\in D)$, with:

$$D=<\mathcal{P}_2,\substack{\circ\\|\\\bullet},\substack{\bullet\\|\\\circ}>=P_2$$

Thus, we are led to the conclusion in the statement. □

There are many other examples of easy groups, and we will gradually explore this. To start with, we have the following interesting result, still in the continuous case:

Theorem 12.8. *We have the following results:*

1. *The unitary bistochastic group C_N is easy, coming from the category $\mathcal{P}_{12}$ of matching singletons and pairings.*
2. *The orthogonal bistochastic group B_N is easy, coming from the category P_{12} of singletons and pairings.*

Proof. The proof here is similar to the proof of Theorem 12.7. To be more precise, we can use the results there, and the proof goes as follows:

(1) The group $C_N \subset U_N$ is defined by imposing the following relations, with ξ being the all-one vector, which correspond to the bistochasticity condition:

$$u\xi = \xi \quad , \quad \bar{u}\xi = \xi$$

But these relations tell us precisely that the following two operators, with the partitions on the right being singletons, must be in the associated Tannakian category C:

$$T_\pi \quad : \quad \pi = \overset{|}{\circ}\ ,\ \overset{|}{\bullet}$$

Thus the associated Tannakian category is $C = span(T_\pi | \pi \in D)$, with:

$$D = < \mathcal{P}_2, \overset{|}{\circ}, \overset{|}{\bullet} > = \mathcal{P}_{12}$$

Thus, we are led to the conclusion in the statement.

(2) In order to deal now with the real bistochastic group B_N, we can either use a similar argument, or simply use the following intersection formula:

$$B_N = C_N \cap O_N$$

Indeed, at the categorical level, this intersection formula tells us that the associated Tannakian category is given by $C = span(T_\pi | \pi \in D)$, with:

$$D = < \mathcal{P}_{12}, P_2 > = P_{12}$$

Thus, we are led to the conclusion in the statement. □

As a comment here, we have used in the above the fact, which is something quite trivial, that the category of partitions associated to an intersection of easy quantum groups is generated by the corresponding categories of partitions. We will be back to this, and to some other product operations as well, with similar results, later on.

We can put now the results that we have together, as follows:

Theorem 12.9. *The basic unitary and bistochastic groups,*

$$\begin{tikzcd}
C_N \arrow[r] & U_N \\
B_N \arrow[u] \arrow[r] & O_N \arrow[u]
\end{tikzcd}$$

are all easy, coming from the various categories of singletons and pairings.

Proof. We know from the above that the groups in the statement are indeed easy, the corresponding diagram of categories of partitions being as follows:

$$\begin{tikzcd}
\mathcal{P}_{12} \arrow[d] & \mathcal{P}_2 \arrow[l] \arrow[d] \\
P_{12} & P_2 \arrow[l]
\end{tikzcd}$$

Thus, we are led to the conclusion in the statement. □

Summarizing, what we have so far is a general notion of "easiness", coming from the Brauer theorems for O_N, U_N, and their straightforward extensions to B_N, C_N.

12b. Reflection groups

In view of the above, the notion of easiness is a quite interesting one, deserving a full, systematic investigation. As a first natural question that we would like to solve, we would like to compute the easy group associated to the category of all partitions P itself.

And here, no surprise, we are led to the most basic, but non-trivial, classical group that we know, namely the symmetric group S_N. To be more precise, we have the following Brauer type theorem for S_N, which answers our question formulated above:

Theorem 12.10. *The symmetric group S_N, regarded as group of unitary matrices,*

$$S_N \subset O_N \subset U_N$$

via the permutation matrices, is easy, coming from the category of all partitions P.

Proof. Consider indeed the group S_N, regarded as a group of unitary matrices, with each permutation $\sigma \in S_N$ corresponding to the associated permutation matrix:

$$\sigma(e_i) = e_{\sigma(i)}$$

Consider as well the easy group $G\subset O_N$ coming from the category of all partitions P. Since P is generated by the one-block "fork" partition $Y\in P(2,1)$, we have:

$$C(G)=C(O_N)\Big/\Big\langle T_Y\in Hom(u^{\otimes 2},u)\Big\rangle$$

The linear map associated to Y is given by the following formula:

$$T_Y(e_i\otimes e_j)=\delta_{ij}e_i$$

In order to do the computations, we use the following formulae:

$$u=(u_{ij})_{ij}\quad,\quad u^{\otimes 2}=(u_{ij}u_{kl})_{ik,jl}\quad,\quad T_Y=(\delta_{ijk})_{i,jk}$$

We therefore obtain the following formula:

$$(T_Yu^{\otimes 2})_{i,jk}=\sum_{lm}(T_Y)_{i,lm}(u^{\otimes 2})_{lm,jk}=u_{ij}u_{ik}$$

On the other hand, we have as well the following formula:

$$(uT_Y)_{i,jk}=\sum_l u_{il}(T_Y)_{l,jk}=\delta_{jk}u_{ij}$$

Thus, the relation defining $G\subset O_N$ reformulates as follows:

$$T_Y\in Hom(u^{\otimes 2},u)\iff u_{ij}u_{ik}=\delta_{jk}u_{ij},\forall i,j,k$$

In other words, the elements u_{ij} must be projections, which must be pairwise orthogonal on the rows of $u=(u_{ij})$. We conclude that $G\subset O_N$ is the subgroup of matrices $g\in O_N$ having the property $g_{ij}\in\{0,1\}$. Thus we have $G=S_N$, as desired. □

As a continuation of this, let us discuss now the hyperoctahedral group H_N. The result here is quite similar to the one for the symmetric groups, as follows:

Theorem 12.11. *The hyperoctahedral group H_N, regarded as a group of matrices,*

$$S_N\subset H_N\subset O_N$$

is easy, coming from the category of partitions with even blocks P_{even}.

Proof. This follows as usual from Tannakian duality. To be more precise, consider the following one-block partition, which, as the name indicates, looks like a H letter:

$$H\in P(2,2)$$

The linear map associated to this partition is then given by:

$$T_H(e_i\otimes e_j)=\delta_{ij}e_i\otimes e_i$$

By using this formula, we have the following computation:

$$\begin{aligned}(T_H\otimes id)u^{\otimes 2}(e_a\otimes e_b)&=&(T_H\otimes id)\left(\sum_{ijkl}e_{ij}\otimes e_{kl}\otimes u_{ij}u_{kl}\right)(e_a\otimes e_b)\\&=&(T_H\otimes id)\left(\sum_{ik}e_i\otimes e_k\otimes u_{ia}u_{kb}\right)\\&=&\sum_i e_i\otimes e_i\otimes u_{ia}u_{ib}\end{aligned}$$

On the other hand, we have as well the following computation:

$$\begin{aligned}u^{\otimes 2}(T_H\otimes id)(e_a\otimes e_b)&=&\delta_{ab}\left(\sum_{ijkl}e_{ij}\otimes e_{kl}\otimes u_{ij}u_{kl}\right)(e_a\otimes e_a)\\&=&\delta_{ab}\sum_{ij}e_i\otimes e_k\otimes u_{ia}u_{ka}\end{aligned}$$

We conclude from this that we have the following equivalence:

$$T_H\in End(u^{\otimes 2})\iff\delta_{ik}u_{ia}u_{ib}=\delta_{ab}u_{ia}u_{ka},\forall i,k,a,b$$

But the relations on the right tell us that the entries of the matrix $u=(u_{ij})$ must satisfy the following condition, on each row and column of u:

$$\alpha\beta=0$$

We conclude that that the corresponding closed subgroup $G\subset O_N$ consists of the matrices $g\in O_N$ which are permutation-like, with ± 1 nonzero entries. Thus, the corresponding group is $G=H_N$, and as a conclusion to this, we have:

$$C(H_N)=C(O_N)\Big/\Big\langle T_H\in End(u^{\otimes 2})\Big\rangle$$

According now to our conventions for easiness, this means that the hyperoctahedral group H_N is easy, coming from the following category of partitions:

$$D=<H>$$

But the category on the right can be computed by drawing pictures, and we have:

$$<H>=P_{even}$$

Thus, we are led to the conclusion in the statement. □

More generally now, we have in fact the following grand result, regarding the series of complex reflection groups H_N^s, which covers both the groups S_N,H_N:

Theorem 12.12. *The complex reflection group $H_N^s=\mathbb Z_s\wr S_N$ is easy, the corresponding category P^s consisting of the partitions satisfying the condition*

$$\#\circ=\#\bullet\,(s)$$

as a weighted sum, in each block. In particular, we have the following results:

1. *S_N is easy, coming from the category P.*
2. *$H_N=\mathbb Z_2\wr S_N$ is easy, coming from the category P_{even}.*
3. *$K_N=\mathbb T\wr S_N$ is easy, coming from the category $\mathcal P_{even}$.*

Proof. This is something that we already know at $s=1,2$, from Theorems 12.10 and 12.11. In general, the proof is similar, based on Tannakian duality. To be more precise, in what regards the main assertion, the idea here is that the one-block partition $\pi\in P(s)$, which generates the category of partitions P^s in the statement, implements the relations producing the subgroup $H_N^s\subset S_N$. As for the last assertions, these are all elementary:

(1) At $s=1$ we know that we have $H_N^1=S_N$. Regarding now the corresponding category, here the condition $\#\circ=\#\bullet\,(1)$ is automatic, and so $P^1=P$.

(2) At $s=2$ we know that we have $H_N^2=H_N$. Regarding now the corresponding category, here the condition $\#\circ=\#\bullet\,(2)$ reformulates as follows:

$$\#\circ+\#\bullet=0(2)$$

Thus each block must have even size, and we obtain, as claimed, $P^2=P_{even}$.

(3) At $s=\infty$ we know that we have $H_N^\infty=K_N$. Regarding now the corresponding category, here the condition $\#\circ=\#\bullet\,(\infty)$ reads:

$$\#\circ=\#\bullet$$

But this is the condition defining $\mathcal P_{even}$, and so $P^\infty=\mathcal P_{even}$, as claimed. □

Summarizing, we have many examples. In fact, our list of easy groups has currently become quite big, and here is a selection of the main results that we have so far:

Theorem 12.13. *We have a diagram of compact groups as follows,*

$$\begin{tikzcd}
K_N \arrow[r] & U_N \\
H_N \arrow[r] \arrow[u] & O_N \arrow[u]
\end{tikzcd}$$

where $H_N=\mathbb Z_2\wr S_N$ and $K_N=\mathbb T\wr S_N$, and all these groups are easy.

PROOF. This follows from the above results. To be more precise, we know that the above groups are all easy, the corresponding categories of partitions being as follows:

$$\begin{tikzcd}
\mathcal{P}_{even} \arrow[d] & \mathcal{P}_2 \arrow[l] \arrow[d] \\
P_{even} & P_2 \arrow[l]
\end{tikzcd}$$

Thus, we are led to the conclusion in the statement. □

Summarizing, most of the groups that we investigated so far in this book are covered by the easy group formalism. Which is something very nice, and good to know.

As a comment here, one notable exception is the symplectic group Sp_N. But this group is covered in fact as well, by a suitable extension of the easy group formalism.

12c. Basic operations

Let us discuss now some basic composition operations, in general, and for the easy groups. We will be mainly interested in the following operations:

DEFINITION 12.14. *The closed subgroups of U_N are subject to intersection and generation operations, constructed as follows:*

1. *Intersection: $H \cap K$ is the usual intersection of H, K.*
2. *Generation: $< H, K >$ is the closed subgroup generated by H, K.*

Alternatively, we can define these operations at the function algebra level, by performing certain operations on the associated ideals, as follows:

PROPOSITION 12.15. *Assuming that we have presentation results as follows,*

$$C(H) = C(U_N)/I \quad , \quad C(K) = C(U_N)/J$$

the groups $H \cap K$ and $< H, K >$ are given by the following formulae,

$$C(H \cap K) = C(U_N)/ < I, J >$$

$$C(< H, K >) = C(U_N)/(I \cap J)$$

at the level of the associated algebras of functions.

PROOF. This is indeed clear from the definition of the operations $\cap$ and $<, >$, as formulated above, and from the Stone-Weierstrass theorem. □

In what follows we will need Tannakian formulations of the above two operations. The result here, that we have already used a couple of times in the above, is as follows:

THEOREM 12.16. *The intersection and generation operations $\cap$ and $<,>$ can be constructed via the Tannakian correspondence $G \to C_G$, as follows:*

1. *Intersection: defined via $C_{G\cap H} =< C_G, C_H >$.*
2. *Generation: defined via $C_{<G,H>} = C_G \cap C_H$.*

PROOF. This follows from Proposition 12.15, and from Tannakian duality. Indeed, it follows from Tannakian duality that given a closed subgroup $G \subset U_N$, with fundamental representation v, the algebra of functions $C(G)$ has the following presentation:

$$C(G) = C(U_N)\Big/\Big\langle T \in Hom(u^{\otimes k}, u^{\otimes l})\Big|\forall k, \forall l, \forall T \in Hom(v^{\otimes k}, v^{\otimes l})\Big\rangle$$

In other words, given a closed subgroup $G \subset U_N$, we have a presentation of the following type, with I_G being the ideal coming from the Tannakian category of G:

$$C(G) = C(U_N)/I_G$$

But this leads to the conclusion in the statement. □

In relation now with our easiness questions, we first have the following result:

PROPOSITION 12.17. *Assuming that H, K are easy, then so is $H \cap K$, and we have*

$$D_{H\cap K} =< D_H, D_K >$$

at the level of the corresponding categories of partitions.

PROOF. We have indeed the following computation:

$$\begin{aligned} C_{H\cap K} &= < C_H, C_K > \\ &= < span(D_H), span(D_K) > \\ &= span(< D_H, D_K >) \end{aligned}$$

Thus, by Tannakian duality we obtain the result. □

Regarding now the generation operation, the situation here is more complicated, due to a number of technical reasons, and we only have the following statement:

PROPOSITION 12.18. *Assuming that H, K are easy, we have an inclusion*

$$< H, K >\subset \{H, K\}$$

coming from an inclusion of Tannakian categories as follows,

$$C_H \cap C_K \supset span(D_H \cap D_K)$$

where $\{H, K\}$ is the easy group having as category of partitions $D_H \cap D_K$.

Proof. This follows from the definition and properties of the generation operation, explained above, and from the following computation:

$$\begin{aligned} C_{<H,K>} &= C_H \cap C_K \\ &= span(D_H) \cap span(D_K) \\ &\supset span(D_H \cap D_K) \end{aligned}$$

Indeed, by Tannakian duality we obtain from this all the assertions. □

It is not clear if the inclusions in Proposition 12.18 are isomorphisms or not, and this even under a supplementary $N >> 0$ assumption. Technically speaking, the problem comes from the fact that the operation $\pi \to T_\pi$ does not produce linearly independent maps, and so all that we are doing is sensitive to the value of $N \in \mathbb{N}$. The subject here is quite technical, to be further developed in Part IV below, with probabilistic motivations in mind, without however solving the present algebraic questions.

Summarizing, we have some problems here, and we must proceed as follows:

Theorem 12.19. *The intersection and easy generation operations $\cap$ and $\{\, ,\}$ can be constructed via the Tannakian correspondence $G \to D_G$, as follows:*

1. *Intersection: defined via $D_{G\cap H} =< D_G, D_H >$.*
2. *Easy generation: defined via $D_{\{G,H\}} = D_G \cap D_H$.*

Proof. Here the situation is as follows:

(1) This is a true and honest result, coming from Proposition 12.17.

(2) This is more of an empty statement, coming from Proposition 12.18. □

As already mentioned, there is some interesting mathematics still to be worked out, in relation with all this, and we will be back to this later, with further details. With the above notions in hand, however, even if not fully satisfactory, we can formulate a nice result, which improves our main result so far, namely Theorem 12.13, as follows:

Theorem 12.20. *The basic unitary and reflection groups, namely*

$$\begin{tikzcd} K_N \arrow[r] & U_N \\ H_N \arrow[u] \arrow[r] & O_N \arrow[u] \end{tikzcd}$$

are all easy, and they form an intersection and easy generation diagram, in the sense that the above square diagram satisfies $U_N = \{K_N, O_N\}$, and $H_N = K_N \cap O_N$.

Proof. We know from Theorem 12.13 that the groups in the statement are easy, the corresponding categories of partitions being as follows:

$$\begin{tikzcd}
\mathcal{P}_{even} \arrow[d] & \mathcal{P}_2 \arrow[l] \arrow[d] \\
P_{even} & P_2 \arrow[l]
\end{tikzcd}$$

Now observe that this latter diagram is an intersection and generation diagram. By using Theorem 12.19, this reformulates into the fact that the diagram of quantum groups is an intersection and easy generation diagram, as claimed. □

It is possible to further improve the above result, by proving that the diagram there is actually a plain generation diagram. However, this is something more technical, and for a discussion here, you can check for instance my quantum group book [11].

Moving forward, as a continuation of the above, it is possible to develop some more general theory, along the above lines. Given a closed subgroup $G \subset U_N$, we can talk about its "easy envelope", which is the smallest easy group $\widetilde{G}$ containing G. This easy envelope appears by definition as an intermediate closed subgroup, as follows:

$$G \subset \widetilde{G} \subset U_N$$

With this notion in hand, Proposition 12.18 can be refined into a result stating that given two easy groups H, K, we have inclusions as follows:

$$< H, K > \subset < \widetilde{H, K} > \subset \{H, K\}$$

In order to discuss all this, let us start with the following definition:

Definition 12.21. *A closed subgroup $G \subset U_N$ is called homogeneous when*

$$S_N \subset G \subset U_N$$

with $S_N \subset U_N$ being the standard embedding, via permutation matrices.

We will be interested in such groups, which cover for instance all the easy groups, and many more. At the Tannakian level, we have the following result:

Theorem 12.22. *The homogeneous groups $S_N \subset G \subset U_N$ are in one-to-one correspondence with the intermediate tensor categories*

$$span\left(T_\pi\Big|\pi \in \mathcal{P}_2\right) \subset C \subset span\left(T_\pi\Big|\pi \in P\right)$$

where P is the category of all partitions, $\mathcal{P}_2$ is the category of the matching pairings, and $\pi \to T_\pi$ is the standard implementation of partitions, as linear maps.

Proof. This follows from Tannakian duality, and from the Brauer type results for S_N, U_N. To be more precise, we know from Tannakian duality that each closed subgroup $G \subset U_N$ can be reconstructed from its Tannakian category $C = (C(k,l))$, as follows:

$$C(G) = C(U_N) \Big/ \left\langle T \in Hom(u^{\otimes k}, u^{\otimes l}) \Big| \forall k, l, \forall T \in C(k,l) \right\rangle$$

Thus we have a one-to-one correspondence $G \leftrightarrow C$, given by Tannakian duality, and since the endpoints $G = S_N, U_N$ are both easy, corresponding to the categories $C = span(T_\pi | \pi \in D)$ with $D = P, \mathcal{P}_2$, this gives the result. □

Our purpose now will be that of using the Tannakian result in Theorem 12.22, in order to introduce and study a combinatorial notion of "easiness level", for the arbitrary intermediate groups $S_N \subset G \subset U_N$. Let us begin with the following simple fact:

Proposition 12.23. *Given a homogeneous group $S_N \subset G \subset U_N$, with associated Tannakian category $C = (C(k,l))$, the sets*

$$D^1(k,l) = \left\{ \pi \in P(k,l) \Big| T_\pi \in C(k,l) \right\}$$

form a category of partitions, in the sense of Definition 12.3.

Proof. We use the basic categorical properties of the correspondence $\pi \to T_\pi$ between partitions and linear maps, that we established in the above, namely:

$$T_{[\pi\sigma]} = T_\pi \otimes T_\sigma \quad , \quad T_{[^\sigma_\pi]} \sim T_\pi T_\sigma \quad , \quad T_{\pi^*} = T_\pi^*$$

Together with the fact that C is a tensor category, we deduce from these formulae that we have the following implication:

$$\begin{aligned} \pi, \sigma \in D^1 &\implies T_\pi, T_\sigma \in C \\ &\implies T_\pi \otimes T_\sigma \in C \\ &\implies T_{[\pi\sigma]} \in C \\ &\implies [\pi\sigma] \in D^1 \end{aligned}$$

On the other hand, we have as well the following implication:

$$\begin{aligned} \pi, \sigma \in D^1 &\implies T_\pi, T_\sigma \in C \\ &\implies T_\pi T_\sigma \in C \\ &\implies T_{[^\sigma_\pi]} \in C \\ &\implies [^\sigma_\pi] \in D^1 \end{aligned}$$

Finally, we have as well the following implication:

$$\begin{aligned}\pi \in D^1 &\implies T_\pi \in C\\ &\implies T_\pi^* \in C\\ &\implies T_{\pi^*} \in C\\ &\implies \pi^* \in D^1\end{aligned}$$

Thus D^1 is indeed a category of partitions, as claimed. □

We can further refine the above observation, in the following way:

Proposition 12.24. *Given a compact group $S_N \subset G \subset U_N$, construct $D^1 \subset P$ as above, and let $S_N \subset G^1 \subset U_N$ be the easy group associated to D^1. Then:*

1. *We have $G \subset G^1$, as subgroups of U_N.*
2. *G^1 is the smallest easy group containing G.*
3. *G is easy precisely when $G \subset G^1$ is an isomorphism.*

Proof. All this is elementary, the proofs being as follows:

(1) We know that the Tannakian category of G^1 is given by:

$$C^1_{kl} = span\left(T_\pi\Big|\pi \in D^1(k,l)\right)$$

Thus we have $C^1 \subset C$, and so $G \subset G^1$, as subgroups of U_N.

(2) Assuming that we have $G \subset G'$, with G' easy, coming from a Tannakian category $C' = span(D')$, we must have $C' \subset C$, and so $D' \subset D^1$. Thus, $G^1 \subset G'$, as desired.

(3) This is a trivial consequence of (2). □

Summarizing, we have now a notion of "easy envelope", as follows:

Definition 12.25. *The easy envelope of a homogeneous group $S_N \subset G \subset U_N$ is the easy group $S_N \subset G^1 \subset U_N$ associated to the category of partitions*

$$D^1(k,l) = \left\{\pi \in P(k,l)\Big|T_\pi \in C(k,l)\right\}$$

where $C = (C(k,l))$ is the Tannakian category of G.

At the level of examples, most of the known homogeneous groups $S_N \subset G \subset U_N$ are in fact easy. However, there are non-easy interesting examples as well, such as the generic reflection groups H_N^{sd} from chapter 10, and we will certainly have an exercise at the end of this chapter, regarding the computation of the corresponding easy envelopes.

As a technical observation now, we can in fact generalize the above construction to any closed subgroup $G \subset U_N$, and we have the following result:

Proposition 12.26. *Given a closed subgroup $G\subset U_N$, construct $D^1\subset P$ as above, and let $S_N\subset G^1\subset U_N$ be the easy group associated to D^1. We have then*

$$G^1=(<G,S_N>)^1$$

where $<G,S_N>\subset U_N$ is the smallest closed subgroup containing G,S_N.

Proof. According to our Tannakian results, the subgroup $<G,S_N>\subset U_N$ in the statement exists indeed, and can be obtained by intersecting categories, as follows:

$$C_{<G,S_N>}=C_G\cap C_{S_N}$$

We conclude from this that for any $\pi\in P(k,l)$ we have:

$$T_\pi\in C_{<G,S_N>}(k,l)\iff T_\pi\in C_G(k,l)$$

It follows that the D^1 categories for the groups $<G,S_N>$ and G coincide, and so the easy envelopes $(<G,S_N>)^1$ and G^1 coincide as well, as stated. □

In order now to fine-tune all this, by using an arbitrary parameter $p\in\mathbb{N}$, which can be thought of as being an "easiness level", we can proceed as follows:

Definition 12.27. *Given a compact group $S_N\subset G\subset U_N$, and an integer $p\in\mathbb{N}$, we construct the family of linear spaces*

$$E^p(k,l)=\left\{\alpha_1T_{\pi_1}+\ldots+\alpha_pT_{\pi_p}\in C(k,l)\Big|\alpha_i\in\mathbb{C},\pi_i\in P(k,l)\right\}$$

and we denote by C^p the smallest tensor category containing $E^p=(E^p(k,l))$, and by $S_N\subset G^p\subset U_N$ the compact group corresponding to this category C^p.

As a first observation, at $p=1$ we have $C^1=E^1=span(D^1)$, where D^1 is the category of partitions constructed in Proposition 12.24. Thus the group G^1 constructed above coincides with the "easy envelope" of G, from Definition 12.25.

In the general case, $p\in\mathbb{N}$, the family $E^p=(E^p(k,l))$ constructed above is not necessarily a tensor category, but we can of course consider the tensor category C^p generated by it, as indicated. Finally, in the above definition we have used of course the Tannakian duality results, in order to perform the operation $C^p\to G^p$.

In practice, the construction in Definition 12.27 is often something quite complicated, and it is convenient to use the following observation:

Proposition 12.28. *The category C^p constructed above is generated by the spaces*

$$E^p(l)=\left\{\alpha_1T_{\pi_1}+\ldots+\alpha_pT_{\pi_p}\in C(l)\Big|\alpha_i\in\mathbb{C},\pi_i\in P(l)\right\}$$

where $C(l)=C(0,l),P(l)=P(0,l)$, with l ranging over the colored integers.

Proof. We use the well-known fact, that we know from chapter 11, that given a closed subgroup $G\subset U_N$, we have a Frobenius type isomorphism, as follows:

$$Hom(u^{\otimes k},u^{\otimes l})\simeq Fix(u^{\otimes\bar{k}l})$$

If we apply this to the group G^p, we obtain an isomorphism as follows:

$$C(k,l)\simeq C(\bar{k}l)$$

On the other hand, we have as well an isomorphism $P(k,l)\simeq P(\bar{k}l)$, obtained by performing a counterclockwise rotation to the partitions $\pi\in P(k,l)$. According to the above definition of the spaces $E^p(k,l)$, this induces an isomorphism as follows:

$$E^p(k,l)\simeq E^p(\bar{k}l)$$

We deduce from this that for any partitions $\pi_1,\ldots,\pi_p\in C(k,l)$, having rotated versions $\rho_1,\ldots,\rho_p\in C(\bar{k}l)$, and for any scalars $\alpha_1,\ldots,\alpha_p\in\mathbb{C}$, we have:

$$\alpha_1T_{\pi_1}+\ldots+\alpha_pT_{\pi_p}\in C(k,l)\iff\alpha_1T_{\rho_1}+\ldots+\alpha_pT_{\rho_p}\in C(\bar{k}l)$$

But this gives the conclusion in the statement, and we are done. □

The main properties of the construction $G\to G^p$ can be summarized as follows:

Theorem 12.29. *Given a compact group $S_N\subset G\subset U_N$, the compact groups G^p constructed above form a decreasing family, whose intersection is G:*

$$G=\bigcap_{p\in\mathbb{N}}G^p$$

Moreover, G is easy when this decreasing limit is stationary, $G=G^1$.

Proof. By definition of $E^p(k,l)$, and by using Proposition 12.28, these linear spaces form an increasing filtration of $C(k,l)$. The same remains true when completing into tensor categories, and so we have an increasing filtration, as follows:

$$C=\bigcup_{p\in\mathbb{N}}C^p$$

At the compact group level now, we obtain the decreasing intersection in the statement. Finally, the last assertion is clear from Proposition 12.28. □

As a main consequence of the above results, we can now formulate:

Definition 12.30. *We say that a homogeneous compact group*

$$S_N\subset G\subset U_N$$

is easy at order p when $G=G^p$, with p being chosen minimal with this property.

Observe that the order 1 notion corresponds to the usual easiness. In general, all this is quite abstract, but there are several explicit examples, that can be worked out. For more on all this, you can check my quantum group book [11].

12d. Classification results

Let us go back now to plain easiness, and discuss some classification results, following the old papers, and then the more recent paper of Tarrago-Weber. In order to cut from the complexity, we must impose an extra axiom, and we will use here:

THEOREM 12.31. *For an easy group $G = (G_N)$, coming from a category of partitions $D \subset P$, the following conditions are equivalent:*

1. *$G_{N-1} = G_N \cap U_{N-1}$, via the embedding $U_{N-1} \subset U_N$ given by $u \to diag(u, 1)$.*
2. *$G_{N-1} = G_N \cap U_{N-1}$, via the N possible diagonal embeddings $U_{N-1} \subset U_N$.*
3. *D is stable under the operation which consists in removing blocks.*

If these conditions are satisfied, we say that $G = (G_N)$ is uniform.

PROOF. We use the general easiness theory explained above, as follows:

(1) $\Longleftrightarrow$ (2) This is something standard, coming from the inclusion $S_N \subset G_N$, which makes everything S_N-invariant. The result follows as well from the proof of (1) $\Longleftrightarrow$ (3) below, which can be converted into a proof of (2) $\Longleftrightarrow$ (3), in the obvious way.

(1) $\Longleftrightarrow$ (3) Given a subgroup $K \subset U_{N-1}$, with fundamental representation u, consider the $N \times N$ matrix $v = diag(u, 1)$. Our claim is that for any $\pi \in P(k)$ we have:

$$\xi_\pi \in Fix(v^{\otimes k}) \iff \xi_{\pi'} \in Fix(v^{\otimes k'}), \forall \pi' \in P(k'), \pi' \subset \pi$$

In order to prove this, we must study the condition on the left. We have:

$$\begin{aligned}
\xi_\pi \in Fix(v^{\otimes k}) &\iff (v^{\otimes k}\xi_\pi)_{i_1\ldots i_k} = (\xi_\pi)_{i_1\ldots i_k}, \forall i\\
&\iff \sum_j (v^{\otimes k})_{i_1\ldots i_k, j_1\ldots j_k}(\xi_\pi)_{j_1\ldots j_k} = (\xi_\pi)_{i_1\ldots i_k}, \forall i\\
&\iff \sum_j \delta_\pi(j_1, \ldots, j_k) v_{i_1 j_1} \ldots v_{i_k j_k} = \delta_\pi(i_1, \ldots, i_k), \forall i
\end{aligned}$$

Now let us recall that our representation has the special form $v = diag(u, 1)$. We conclude from this that for any index $a \in \{1, \ldots, k\}$, we must have:

$$i_a = N \implies j_a = N$$

With this observation in hand, if we denote by i', j' the multi-indices obtained from i, j obtained by erasing all the above $i_a = j_a = N$ values, and by $k' \leq k$ the common length of these new multi-indices, our condition becomes:

$$\sum_{j'} \delta_\pi(j_1, \ldots, j_k)(v^{\otimes k'})_{i'j'} = \delta_\pi(i_1, \ldots, i_k), \forall i$$

Here the index j is by definition obtained from j' by filling with N values. In order to finish now, we have two cases, depending on i, as follows:

Case 1. Assume that the index set $\{a|i_a = N\}$ corresponds to a certain subpartition $\pi' \subset \pi$. In this case, the N values will not matter, and our formula becomes:

$$\sum_{j'} \delta_\pi(j'_1, \ldots, j'_{k'})(v^{\otimes k'})_{i'j'} = \delta_\pi(i'_1, \ldots, i'_{k'})$$

Case 2. Assume now the opposite, namely that the set $\{a|i_a = N\}$ does not correspond to a subpartition $\pi' \subset \pi$. In this case the indices mix, and our formula reads:

$$0 = 0$$

Thus, we are led to $\xi_{\pi'} \in Fix(v^{\otimes k'})$, for any subpartition $\pi' \subset \pi$, as claimed.

Now with this claim in hand, the result follows from Tannakian duality. □

We can now formulate a first classification result, as follows:

Theorem 12.32. *The uniform orthogonal easy groups are as follows,*

$$\begin{tikzcd} B_N \arrow[r] & O_N \\ S_N \arrow[u] \arrow[r] & H_N \arrow[u] \end{tikzcd}$$

and this diagram is an intersection and easy generation diagram.

Proof. We know that the quantum groups in the statement are indeed easy and uniform, the corresponding categories of partitions being as follows:

$$\begin{tikzcd} P_{12} \arrow[d] & P_2 \arrow[l] \arrow[d] \\ P & P_{even} \arrow[l] \end{tikzcd}$$

Since this latter diagram is an intersection and generation diagram, we conclude that we have an intersection and easy generation diagram of quantum groups, as stated. Regarding now the classification, consider an arbitrary easy group, as follows:

$$S_N \subset G_N \subset O_N$$

This group must then come from a category of partitions, as follows:

$$P_2 \subset D \subset P$$

Now if we assume $G = (G_N)$ to be uniform, this category of partitions D is uniquely determined by the subset $L \subset \mathbb{N}$ consisting of the sizes of the blocks of the partitions in D. Our claim now is that the admissible sets are as follows:

(1) $L = \{2\}$, producing O_N.

(2) $L = \{1, 2\}$, producing B_N.

(3) $L = \{2, 4, 6, \ldots\}$, producing H_N.

(4) $L = \{1, 2, 3, \ldots\}$, producing S_N.

Indeed, in one sense, this follows from our easiness results for O_N, B_N, H_N, S_N. In the other sense now, assume that $L \subset \mathbb{N}$ is such that the set P_L consisting of partitions whose sizes of the blocks belong to L is a category of partitions. We know from the axioms of the categories of partitions that the semicircle $\cap$ must be in the category, so we have $2 \in L$. Our claim is that the following conditions must be satisfied as well:

$$k, l \in L,\ k > l \implies k - l \in L$$

$$k \in L,\ k \geq 2 \implies 2k - 2 \in L$$

Indeed, we will prove that both conditions follow from the axioms of the categories of partitions. Let us denote by $b_k \in P(0, k)$ the one-block partition, as follows:

$$b_k = \begin{Bmatrix} \sqcap\!\sqcap & \ldots & \sqcap \\ 1\ \ 2 & \ldots & k \end{Bmatrix}$$

For $k > l$, we can write b_{k-l} in the following way:

$$b_{k-l} = \left\{\begin{matrix} \sqcap\!\sqcap & \ldots & \ldots & \ldots & \ldots & \sqcap \\ 1\ \ 2 & \ldots & l & l+1 & \ldots & k \\ \sqcup\!\sqcup & \ldots & \sqcup & | & \ldots & | \\ & & & 1 & \ldots & k-l \end{matrix}\right\}$$

In other words, we have the following formula:

$$b_{k-l} = (b_l^* \otimes |^{\otimes k-l}) b_k$$

Since all the terms of this composition are in P_L, we have $b_{k-l} \in P_L$, and this proves our first formula. As for the second formula, this can be proved in a similar way, by capping two adjacent k-blocks with a 2-block, in the middle.

With the above two formulae in hand, we can conclude in the following way:

Case 1. Assume $1 \in L$. By using the first formula with $l = 1$ we get:

$$k \in L \implies k - 1 \in L$$

This condition shows that we must have $L = \{1, 2, \ldots, m\}$, for a certain number $m \in \{1, 2, \ldots, \infty\}$. On the other hand, by using the second formula we get:

$$\begin{aligned} m \in L &\implies 2m - 2 \in L \\ &\implies 2m - 2 \leq m \\ &\implies m \in \{1, 2, \infty\} \end{aligned}$$

The case $m = 1$ being excluded by the condition $2 \in L$, we reach to one of the two sets producing the groups S_N, B_N.

Case 2. Assume $1 \notin L$. By using the first formula with $l = 2$ we get:

$$k \in L \implies k - 2 \in L$$

This condition shows that we must have $L = \{2, 4, \ldots, 2p\}$, for a certain number $p \in \{1, 2, \ldots, \infty\}$. On the other hand, by using the second formula we get:

$$\begin{aligned} 2p \in L &\implies 4p - 2 \in L \\ &\implies 4p - 2 \leq 2p \\ &\implies p \in \{1, \infty\} \end{aligned}$$

Thus L must be one of the two sets producing O_N, H_N, and we are done. □

All the above is very nice, but the continuation of the story is more complicated. When lifting the uniformity assumption, the final classification results become more technical, due to the presence of various copies of $\mathbb{Z}_2$, that can be added, while keeping the easiness property still true. To be more precise, in the real case it is known that we have exactly 6 solutions, which are as follows, with the convention $G'_N = G_N \times \mathbb{Z}_2$:

$$\begin{tikzcd}
B_N \arrow[r] & B'_N \arrow[r] & O_N \\
S_N \arrow[u] \arrow[r] & S'_N \arrow[u] \arrow[r] & H_N \arrow[u]
\end{tikzcd}$$

In the unitary case now, the classification is quite similar, but more complicated. In particular the uniform easy groups which are purely unitary are as follows:

$$\begin{tikzcd}
C_N \arrow[r] & U_N \\
S_N \arrow[u] \arrow[r] & K_N \arrow[u]
\end{tikzcd}$$

We refer here to the literature on the subject. Switching topics now, in the projective case, that we are mainly interested in, let us formulate the following definition:

Definition 12.33. *A projective category of pairings is a collection of subsets*

$$NC_2(2k, 2l) \subset E(k, l) \subset P_2(2k, 2l)$$

stable under the usual categorical operations, and satisfying $\sigma \in E \implies |\sigma| \in E$.

As basic examples, going beyond our crossing category setting, we have the following projective categories of pairings, where P_2^* is the category of matching pairings:

$$NC_2\subset P_2^*\subset P_2$$

This follows indeed from definitions. Now with the above notion in hand, we can formulate the following projective analogue of the notion of easiness:

Definition 12.34. *An intermediate compact group*

$$PO_N\subset H\subset PU_N$$

is called projectively easy when its Tannakian category

$$span(NC_2(2k,2l))\subset Hom(v^{\otimes k},v^{\otimes l})\subset span(P_2(2k,2l))$$

comes via via the following formula, using the standard $\pi\to T_\pi$ construction,

$$Hom(v^{\otimes k},v^{\otimes l})=span(E(k,l))$$

for a certain projective category of pairings $E=(E(k,l))$.

Thus, we have a projective notion of easiness, and as examples, the projective versions of easy groups are projectively easy. We have in fact the following general result:

Theorem 12.35. *We have a bijective correspondence between the affine and projective categories of partitions, given by the operation*

$$G\to PG$$

at the level of the corresponding affine and projective easy quantum groups.

Proof. The construction of correspondence $D\to E$ is clear, simply by setting:

$$E(k,l)=D(2k,2l)$$

Indeed, due to the axioms in Definition 12.33, the conditions for categories of partitions are satisfied. Conversely, given $E=(E(k,l))$ as in Definition 12.33, we can set:

$$D(k,l)=\begin{cases}E(k,l)&(k,l\text{ even})\\ \{\sigma:|\sigma\in E(k+1,l+1)\}&(k,l\text{ odd})\end{cases}$$

Our claim is that $D=(D(k,l))$ is a category of partitions. Indeed:

(1) The composition action is clear. Indeed, when looking at the numbers of legs involved, in the even case this is clear, and in the odd case, this follows from:

$$\begin{aligned}|\sigma,|\sigma'\in E\quad&\implies\quad |^\sigma_\tau\in E\\ &\implies\quad {}^\sigma_\tau\in D\end{aligned}$$

(2) For the tensor product axiom, we have 4 cases to be investigated, depending on the parity of the number of legs of σ,τ, as follows:

– The even/even case is clear.

– The odd/even case follows from the following computation:

$$\begin{aligned}|\sigma,\tau\in E &\implies |\sigma\tau\in E\\ &\implies \sigma\tau\in D\end{aligned}$$

– Regarding now the even/odd case, this can be solved as follows:

$$\begin{aligned}\sigma,|\tau\in E &\implies |\sigma|,|\tau\in E\\ &\implies |\sigma||\tau\in E\\ &\implies |\sigma\tau\in E\\ &\implies \sigma\tau\in D\end{aligned}$$

– As for the remaining odd/odd case, here the computation is as follows:

$$\begin{aligned}|\sigma,|\tau\in E &\implies ||\sigma|,|\tau\in E\\ &\implies ||\sigma||\tau\in E\\ &\implies \sigma\tau\in E\\ &\implies \sigma\tau\in D\end{aligned}$$

(3) Finally, the conjugation axiom is clear from definitions. It is also clear that both compositions $D\to E\to D$ and $E\to D\to E$ are the identities, as claimed. As for the quantum group assertion, this is clear as well from definitions. □

12e. Exercises

Exercises:

Exercise 12.36.

Exercise 12.37.

Exercise 12.38.

Exercise 12.39.

Exercise 12.40.

Exercise 12.41.

Exercise 12.42.

Exercise 12.43.

Bonus exercise.

Part IV

Analytic aspects

C'est la vie
Say the old folks
It goes to show
You never can tell

CHAPTER 13

Smooth structure

13a. Differential geometry

Differential geometry.

13b. Smooth structure

Smooth structure.

13c. Standard calculus

Standard calculus.

13d. The complex case

The complex case.

13e. Exercises

Exercises:

Exercise 13.1.

Exercise 13.2.

Exercise 13.3.

Exercise 13.4.

Exercise 13.5.

Exercise 13.6.

Exercise 13.7.

Exercise 13.8.

Bonus exercise.

CHAPTER 14

Metric aspects

14a. Distances, metric

Distances, metric.

14b. Exponential maps

Exponential maps.

14c. Advanced calculus

Advanced calculus.

14d. Complex manifolds

Complex manifolds.

14e. Exercises

Exercises:

Exercise 14.1.

Exercise 14.2.

Exercise 14.3.

Exercise 14.4.

Exercise 14.5.

Exercise 14.6.

Exercise 14.7.

Exercise 14.8.

Bonus exercise.

CHAPTER 15

Integration theory

15a. Uniform integration

Uniform integration.

15b. Weingarten formula

In order to integrate over projective spaces, and other manifolds, let us first investigate the group case. Let us start with the following result, coming from Peter-Weyl:

Theorem 15.1. *For an easy group $G=(G_N)$, coming from a category of partitions $D=(D(k,l))$, the asymptotic moments of the main character are given by*

$$\lim_{N\to\infty}\int_{G_N}\chi^k=\#D(k)$$

where $D(k)=D(\emptyset,k)$, with the limiting sequence on the left consisting of certain integers, and being stationary at least starting from the k-th term.

Proof. This follows indeed from the Peter-Weyl theory, by using the linear independence result for the vectors ξ_π, coming from the Lindstöm determinant formula. □

With these preliminaries in hand, we can now state and prove:

Theorem 15.2. *In the $N\to\infty$ limit, the laws of the main character for the main easy groups, real and complex, and discrete and continuous, are as follows,*

$$\begin{tikzcd}
K_N \arrow[r] & U_N & & B_1 \arrow[r] & G_1 \\
& & : & & \\
H_N \arrow[uu] \arrow[r] & O_N \arrow[uu] & & b_1 \arrow[uu] \arrow[r] & g_1 \arrow[uu]
\end{tikzcd}$$

with these laws, namely the real and complex Gaussian and Bessel laws, being the main limiting laws in real and complex, and discrete and continuous probability.

Proof. This follows from the above results. To be more precise, we know that the above groups are all easy, the corresponding categories of partitions being as follows:

$$\begin{tikzcd}
\mathcal{P}_{even} \arrow[d] & \mathcal{P}_2 \arrow[l] \arrow[d] \\
P_{even} & P_2 \arrow[l]
\end{tikzcd}$$

Thus, we can use Theorem 15.1, are we are led into counting partitions, and then recovering the measures via their moments, and this leads to the result. □

Next, we have the following general formula, also coming from Peter-Weyl:

Theorem 15.3. *The Haar integration over a closed subgroup $G \subset_v U_N$ is given on the dense subalgebra of smooth functions by the Weingarten type formula*

$$\int_G g_{i_1j_1}^{e_1}\ldots g_{i_kj_k}^{e_k}\,dg=\sum_{\pi,\nu\in D(k)}\delta_\pi(i)\delta_\sigma(j)W_k(\pi,\nu)$$

valid for any colored integer $k=e_1\ldots e_k$ and any multi-indices i,j, where $D(k)$ is a linear basis of $Fix(v^{\otimes k})$, the associated generalized Kronecker symbols are given by

$$\delta_\pi(i)=<\pi,e_{i_1}\otimes\ldots\otimes e_{i_k}>$$

and $W_k=G_k^{-1}$ is the inverse of the Gram matrix, $G_k(\pi,\nu)=<\pi,\nu>$.

Proof. This is something very standard, coming from the fact that the above integrals form altogether the orthogonal projection P^k onto the following space:

$$Fix(v^{\otimes k})=span(D(k))$$

Consider now the following linear map, with $D(k)=\{\xi_k\}$ being as in the statement:

$$E(x)=\sum_{\pi\in D(k)}<x,\xi_\pi>\xi_\pi$$

By a standard linear algebra computation, it follows that we have $P=WE$, where W is the inverse of the restriction of E to the following space:

$$K=span\left(T_\pi\Big|\pi\in D(k)\right)$$

But this restriction is the linear map given by the matrix G_k, and so W is the linear map given by the inverse matrix $W_k=G_k^{-1}$, and this gives the result. □

In the easy case, we have the following more concrete result:

Theorem 15.4. *For an easy group $G\subset U_N$, coming from a category of partitions $D=(D(k,l))$, we have the Weingarten formula*

$$\int_G g_{i_1j_1}^{e_1}\ldots g_{i_kj_k}^{e_k}\,dg=\sum_{\pi,\nu\in D(k)}\delta_\pi(i)\delta_\nu(j)W_{kN}(\pi,\nu)$$

for any $k=e_1\ldots e_k$ and any i,j, where $D(k)=D(\emptyset,k)$, δ are usual Kronecker type symbols, checking whether the indices match, and $W_{kN}=G_{kN}^{-1}$, with

$$G_{kN}(\pi,\nu)=N^{|\pi\vee\nu|}$$

where $|.|$ is the number of blocks.

Proof. We use the abstract Weingarten formula, from Theorem 15.3. Indeed, the Kronecker type symbols there are then the usual ones, as shown by:

$$\begin{aligned}
\delta_{\xi_\pi}(i)&=&<\xi_\pi,e_{i_1}\otimes\ldots\otimes e_{i_k}>\\
&=&\left\langle\sum_j\delta_\pi(j_1,\ldots,j_k)e_{j_1}\otimes\ldots\otimes e_{j_k},e_{i_1}\otimes\ldots\otimes e_{i_k}\right\rangle\\
&=&\delta_\pi(i_1,\ldots,i_k)
\end{aligned}$$

The Gram matrix being as well the correct one, we obtain the result. □

Let us go back now to the general easy groups $G\subset U_N$, with the idea in mind of computing the laws of truncated characters. First, we have the following formula:

Proposition 15.5. *The moments of truncated characters are given by the formula*

$$\int_G(g_{11}+\ldots+g_{ss})^kdg=Tr(W_{kN}G_{ks})$$

where G_{kN} and $W_{kN}=G_{kN}^{-1}$ are the associated Gram and Weingarten matrices.

Proof. We have indeed the following computation:

$$\begin{aligned}
\int_G(g_{11}+\ldots+g_{ss})^kdg&=&\sum_{i_1=1}^s\ldots\sum_{i_k=1}^s\int_G g_{i_1i_1}\ldots g_{i_ki_k}\,dg\\
&=&\sum_{\pi,\nu\in D(k)}W_{kN}(\pi,\nu)\sum_{i_1=1}^s\ldots\sum_{i_k=1}^s\delta_\pi(i)\delta_\nu(i)\\
&=&\sum_{\pi,\nu\in D(k)}W_{kN}(\pi,\nu)G_{ks}(\nu,\pi)\\
&=&Tr(W_{kN}G_{ks})
\end{aligned}$$

Thus, we have reached to the formula in the statement. □

In order to process now the above formula, and reach to concrete results, we must impose on our group a uniformity condition. Let us start with:

Proposition 15.6. *For an easy group $G = (G_N)$, coming from a category of partitions $D \subset P$, the following conditions are equivalent:*

1. *$G_{N-1} = G_N \cap U_{N-1}$, via the embedding $U_{N-1} \subset U_N$ given by $u \to diag(u, 1)$.*
2. *$G_{N-1} = G_N \cap U_{N-1}$, via the N possible diagonal embeddings $U_{N-1} \subset U_N$.*
3. *D is stable under the operation which consists in removing blocks.*

If these conditions are satisfied, we say that $G = (G_N)$ is uniform.

Proof. The equivalence $(1) \iff (2)$ comes from the inclusion $S_N \subset G_N$, which makes everything S_N-invariant. Regarding $(1) \iff (3)$, given a subgroup $K \subset_v U_{N-1}$, consider the matrix $u = diag(v, 1)$. Our claim is that for any $\pi \in P(k)$ we have:

$$\xi_\pi \in Fix(u^{\otimes k}) \iff \xi_{\pi'} \in Fix(u^{\otimes k'}), \forall \pi' \in P(k'), \pi' \subset \pi$$

In order to prove this claim, we must study the condition on the left. We have:

$$\begin{aligned} \xi_\pi \in Fix(v^{\otimes k}) &\iff (u^{\otimes k}\xi_\pi)_{i_1 \ldots i_k} = (\xi_\pi)_{i_1 \ldots i_k}, \forall i \\ &\iff \sum_j (u^{\otimes k})_{i_1 \ldots i_k, j_1 \ldots j_k} (\xi_\pi)_{j_1 \ldots j_k} = (\xi_\pi)_{i_1 \ldots i_k}, \forall i \\ &\iff \sum_j \delta_\pi(j_1, \ldots, j_k) u_{i_1 j_1} \ldots u_{i_k j_k} = \delta_\pi(i_1, \ldots, i_k), \forall i \end{aligned}$$

Now let us recall that our representation has the special form $u = diag(v, 1)$. We conclude from this that for any index $a \in \{1, \ldots, k\}$, we have:

$$i_a = N \implies j_a = N$$

With this observation in hand, if we denote by i', j' the multi-indices obtained from i, j obtained by erasing all the above $i_a = j_a = N$ values, and by $k' \leq k$ the common length of these new multi-indices, our condition becomes:

$$\sum_{j'} \delta_\pi(j_1, \ldots, j_k)(u^{\otimes k'})_{i'j'} = \delta_\pi(i_1, \ldots, i_k), \forall i$$

Here the index j is by definition obtained from the index j' by filling with N values. In order to finish now, we have two cases, depending on i, as follows:

<u>Case 1</u>. Assume that the index set $\{a | i_a = N\}$ corresponds to a certain subpartition $\pi' \subset \pi$. In this case, the N values will not matter, and our formula becomes:

$$\sum_{j'} \delta_\pi(j'_1, \ldots, j'_{k'})(u^{\otimes k'})_{i'j'} = \delta_\pi(i'_1, \ldots, i'_{k'})$$

<u>Case 2</u>. Assume now the opposite, namely that the set $\{a | i_a = N\}$ does not correspond to a subpartition $\pi' \subset \pi$. In this case the indices mix, and our formula reads $0 = 0$. Thus we have $\xi_{\pi'} \in Fix(u^{\otimes k'})$ in both cases, for any subpartition $\pi' \subset \pi$, as desired. □

Now back to the laws of truncated characters, we have the following result:

Theorem 15.7. *For a uniform easy group $G = (G_N)$, we have the formula*

$$\lim_{N\to\infty}\int_{G_N}\chi_t^k = \sum_{\pi\in D(k)} t^{|\pi|}$$

with $D \subset P$ being the associated category of partitions.

Proof. We use Proposition 15.5. With $s = [tN]$, the formula there becomes:

$$\int_{G_N}\chi_t^k = Tr(W_{kN}G_{k[tN]})$$

The point now is that in the uniform case the Gram matrix, and so the Weingarten matrix too, is asymptotically diagonal. Thus, we obtain the following estimate:

$$\begin{aligned}\int_{G_N}\chi_t^k &\simeq \sum_{\pi\in D(k)} W_{kN}(\pi,\pi)G_{k[tN]}(\pi,\pi)\\ &\simeq \sum_{\pi\in D(k)} N^{-|\pi|}(tN)^{|\pi|}\\ &= \sum_{\pi\in D(k)} t^{|\pi|}\end{aligned}$$

Thus, we are led to the formula in the statement. □

We can now enlarge our collection of truncated character results, and we have:

Theorem 15.8. *With $N \to \infty$, the laws of truncated characters are as follows:*

1. *For O_N we obtain the Gaussian law g_t.*
2. *For U_N we obtain the complex Gaussian law G_t.*
3. *For S_N we obtain the Poisson law p_t.*
4. *For H_N we obtain the Bessel law b_t.*
5. *For H_N^s we obtain the generalized Bessel law b_t^s.*
6. *For K_N we obtain the complex Bessel law B_t.*

Proof. We already know these results at $t = 1$. In the general case, $t > 0$, these follow via some standard combinatorics, from the formula in Theorem 15.7. □

15c. Projective integrals

Projective integrals.

15d. Quotient spaces

Quotient spaces.

15e. Exercises

Exercises:

Exercise 15.9.

Exercise 15.10.

Exercise 15.11.

Exercise 15.12.

Exercise 15.13.

Exercise 15.14.

Exercise 15.15.

Exercise 15.16.

Bonus exercise.

CHAPTER 16

Free geometry

16a. Free tori

Welcome to freeness. We will be interested here in developing free geometry and analysis, with the hope that all this might be related to physics, at very small scales, quarks and below. The idea being very simple, based on old findings of Heisenberg and others, if it is true indeed that the more you zoom down, the more commutativity dissapears, then, logically, if you zoom hard enough, things will become free.

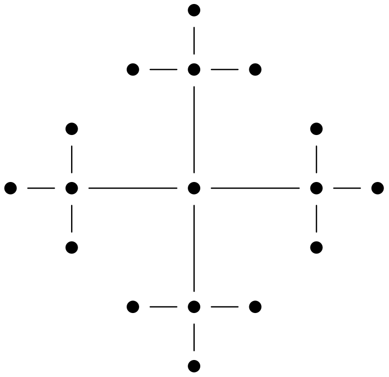
So, what is free? The simplest free object in mathematics is the free group F_N:

Definition 16.1. *The free group F_N is the infinite group*

$$F_N=\left\langle g_1,\ldots,g_N\Big|\,\emptyset\right\rangle$$

generated by N variables $g_1,\ldots,g_N$, with no relations between them.

This might look a bit abstract, but no worries, F_N has some interesting mathematics, coming right away, if you have some knowledge in discrete groups, and know how to look for interesting questions. For instance if you want to draw the Cayley graph of F_N, whose vertices are the elements of F_N, with edges $h-k$ drawn when $h=g_i^{\pm1}k$ for some i, you will end up with an interesting picture, which at $N=2$ looks like this:



And this type of graph certainly has interesting mathematics. One good question for instance is that of computing the number of length $2k$ loops based at the root. Another question, which is in fact equivalent, via moments, is that of computing the Kesten measure of F_N, which is that of the following variable in the group algebra of F_N:

$$\chi = g_1 + \ldots + g_N$$

All this looks very good, we most likely have here our first object of free geometry. In order now to formally understand this, let us recall the following formula, with $\mathbb{T}_N = \mathbb{T}^N$ being the usual torus, and with $\mathbb{Z}^N$ being the free abelian group:

$$\mathbb{T}_N = \widehat{\mathbb{Z}^N}$$

Thus, getting back now to our free group F_N, which is the free analogue of $\mathbb{Z}^N$, it is in fact its dual $\widehat{F_N}$ which is a free manifold, and more specifically the free analogue of $\mathbb{T}^N$. Which is a nice finding, so let us formulate our conclusions as follows:

Definition 16.2. *The free torus $\mathbb{T}_N^+$ is the dual of the free group F_N,*

$$\mathbb{T}_N^+ = \widehat{F_N}$$

in analogy with the fact that the usual torus $\mathbb{T}_N = \mathbb{T}^N$ appears as

$$\mathbb{T}_N = \widehat{\mathbb{Z}^N}$$

with on the right the group $\mathbb{Z}^N$ being the free abelian group.

It is of course possible to formulate things more precisely, and we will be back to this in a moment, but before that, isn't this a bit too abstract? But the point here is that no, at the level of questions to be solved, these remain the same, as for instance the computation of the Kesten measure, which is now a "function" on the free torus:

$$\chi \in C(\mathbb{T}_N^+)$$

In fact, this function is the main character of $\mathbb{T}_N^+$, regarded as a compact quantum group, and so our Kesten problem suddenly becomes something very conceptual, namely the computation of the law of the main character of $\mathbb{T}_N^+$. Which is very nice.

Before getting into details regarding all this, recall that $\mathbb{R}^N$ is as interesting as $\mathbb{C}^N$. So, let us formulate as well the real version of Definition 16.2, as follows:

Definition 16.3. *The free real torus, or free cube, T_N^+ is the dual*

$$T_N^+ = \widehat{L_N}$$

of the group $L_N = F_N / < g_i^2 = 1 >$, in analogy with the fact that the usual cube is

$$T_N = \widehat{\mathbb{Z}_2^N}$$

with on the right the group $\mathbb{Z}_2^N$ being the free real abelian group.

Here the "real" at the end stands for the fact that the generators must satisfy the real reflection condition $g^2 = 1$. As for the fact that "real torus = cube", as stated, this needs some thinking, and in the hope that, after such thinking, you will agree with me that there is indeed a standard torus inside $\mathbb{R}^N$, and that is the unit cube.

Summarizing, all this sounds good, we have a beginning of free geometry, both real and complex, worth developing, by knowing at least what the torus of each theory is. In practice now, at the level of details, in order to talk about $\mathbb{T}_N^+ = \widehat{F_N}$ and $T_N^+ = \widehat{L_N}$ we need an extension of the usual Pontrjagin duality theory for the abelian groups, and this is best done via operator algebras, and the related notion of compact quantum group.

In order to understand all this, let us start with operator algebras. We have:

Definition 16.4. *A C^*-algebra is a complex algebra A, having:*

1. *A norm $a \to ||a||$, making it a Banach algebra.*
2. *An involution $a \to a^*$, satisfying $||aa^*|| = ||a||^2$.*

As basic examples, we have $B(H)$ itself, as well as any norm closed $*$-subalgebra $A \subset B(H)$. It is possible to prove that any C^*-algebra appears in this way, but we will not need in what follows this deep result, called GNS theorem after Gelfand, Naimark, Segal. So, let us simply agree that, by definition, the C^*-algebras A are some sort of "generalized operator algebras", and their elements $a \in A$ can be thought of as being some kind of "generalized operators", on some Hilbert space which is not present.

In practice, this vague idea is all that we need. Indeed, by taking some inspiration from linear algebra, we can emulate spectral theory in our setting, as follows:

Proposition 16.5. *Given $a \in A$, define its spectrum as being the set*

$$\sigma(a) = \left\{\lambda \in \mathbb{C}\Big| a - \lambda \notin A^{-1}\right\}$$

and its spectral radius $\rho(a)$ as the radius of the smallest centered disk containing $\sigma(a)$.

1. *The spectrum of a norm one element is in the unit disk.*
2. *The spectrum of a unitary element ($a^* = a^{-1}$) is on the unit circle.*
3. *The spectrum of a self-adjoint element ($a = a^*$) consists of real numbers.*
4. *The spectral radius of a normal element ($aa^* = a^*a$) is equal to its norm.*

Proof. The first claim is that for any polynomial $f \in \mathbb{C}[X]$, and more generally for any rational function $f \in \mathbb{C}(X)$ having poles outside $\sigma(a)$, we have:

$$\sigma(f(a)) = f(\sigma(a))$$

This indeed something well-known for the usual matrices, and in general, the proof is similar. Regarding now the assertions in the statement, these all follow from this:

(1) This comes from the following formula, valid when $||a||<1$:

$$\frac{1}{1-a}=1+a+a^2+\ldots$$

(2) Assuming $a^*=a^{-1}$, if we denote by D the unit disk, we have, by using (1):

$$||a||=1\implies\sigma(a)\subset D$$

$$||a^{-1}||=1\implies\sigma(a^{-1})\subset D$$

On the other hand, by using the rational function $f(z)=z^{-1}$, we have:

$$\sigma(a^{-1})\subset D\implies\sigma(a)\subset D^{-1}$$

Now by putting everything together we obtain, as desired:

$$\sigma(a)\subset D\cap D^{-1}=\mathbb{T}$$

(3) This follows from (2), by using the rational function $f(z)=(z+it)/(z-it)$. Indeed, for $t>>0$ we have the following computation:

$$\left(\frac{a+it}{a-it}\right)^*=\frac{a-it}{a+it}=\left(\frac{a+it}{a-it}\right)^{-1}$$

Thus the element $f(a)$ is a unitary, and by using (2) its spectrum is contained in $\mathbb{T}$. We conclude from this that we have:

$$f(\sigma(a))=\sigma(f(a))\subset\mathbb{T}$$

But this shows that we have $\sigma(a)\subset f^{-1}(\mathbb{T})=\mathbb{R}$, as desired.

(4) We already know that we have $\rho(a)\leq||a||$, for any $a\in A$. For the reverse inequality, when a is normal, we fix a number $\rho>\rho(a)$. We have then:

$$\begin{aligned}\int_{|z|=\rho}\frac{z^n}{z-a}\,dz&=\int_{|z|=\rho}\sum_{k=0}^\infty z^{n-k-1}a^k\,dz\\&=\sum_{k=0}^\infty\left(\int_{|z|=\rho}z^{n-k-1}dz\right)a^k\\&=a^{n-1}\end{aligned}$$

By applying the norm and taking n-th roots we obtain from this formula:

$$\rho\geq\lim_{n\to\infty}||a^n||^{1/n}$$

When $a=a^*$ we have $||a^n||=||a||^n$ for any exponent of type $n=2^k$, by using the C^*-algebra condition $||aa^*||=||a||^2$, and by taking n-th roots we get, as desired:

$$\rho(a)\geq||a||$$

In the general normal case now, $aa^*=a^*a$, we have $a^n(a^n)^*=(aa^*)^n$, and by using this, along with the result for self-adjoints, applied to aa^*, we obtain:

$$\begin{aligned}\rho(a) &\geq \lim_{n\to\infty}||a^n||^{1/n}\\ &= \sqrt{\lim_{n\to\infty}||a^n(a^n)^*||^{1/n}}\\ &= \sqrt{\lim_{n\to\infty}||(aa^*)^n||^{1/n}}\\ &= \sqrt{\rho(aa^*)}\\ &= ||a||\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Generally speaking, Proposition 16.5 is all you need for doing further operator algebras, only military grade weapons there. As a main application, we have:

Theorem 16.6 (Gelfand). *If X is a compact space, the algebra $C(X)$ of continuous functions $f:X\to\mathbb{C}$ is a commutative C^*-algebra, with structure as follows:*

1. *The norm is the usual sup norm, $||f||=\sup_{x\in X}|f(x)|$.*
2. *The involution is the usual involution, $f^*(x)=\overline{f(x)}$.*

Conversely, any commutative C^*-algebra is of the form $C(X)$, with its "spectrum" $X=Spec(A)$ appearing as the space of characters $\chi:A\to\mathbb{C}$.

Proof. Given a commutative C^*-algebra A, we can define indeed X to be the set of characters $\chi:A\to\mathbb{C}$, with the topology making continuous all the evaluation maps $ev_a:\chi\to\chi(a)$. Then X is a compact space, and $a\to ev_a$ is a morphism of algebras:

$$ev:A\to C(X)$$

We first prove that ev is involutive. We use the following formula:

$$a=\frac{a+a^*}{2}-i\cdot\frac{i(a-a^*)}{2}$$

Thus it is enough to prove the equality $ev_{a^*}=ev_a^*$ for self-adjoint elements a. But this is the same as proving that $a=a^*$ implies that ev_a is a real function, which is in turn true, because $ev_a(\chi)=\chi(a)$ is an element of $\sigma(a)$, contained in $\mathbb{R}$. So, claim proved. Also, since A is commutative, each element is normal, so ev is isometric:

$$||ev_a||=\rho(a)=||a||$$

It remains to prove that ev is surjective. But this follows from the Stone-Weierstrass theorem, because $ev(A)$ is a closed subalgebra of $C(X)$, which separates the points. □

The Gelfand theorem suggests formulating the following definition:

Definition 16.7. *Given a C^*-algebra A, not necessarily commutative, we write*

$$A = C(X)$$

and call the abstract object X a "compact quantum space".

This might look quite revolutionary, but in practice, this definition changes nothing to what we have been doing so far, namely studying the C^*-algebras. So, we will keep studying the C^*-algebras, but by using the above fancy quantum space terminology. For instance whenever we have a morphism $\Phi : A \to B$, we will write $A = C(X), B = C(Y)$, and rather speak of the corresponding morphism $\phi : Y \to X$. And so on.

Now that we have our notion of quantum spaces, good time to get back towards Definitions 16.2 and 16.3. In order to understand what that free tori are, we will need:

Theorem 16.8. *Let Γ be a discrete group, and consider the complex group algebra $\mathbb{C}[\Gamma]$, with involution given by the fact that all group elements are unitaries, $g^* = g^{-1}$.*

(1) *The maximal C^*-seminorm on $\mathbb{C}[\Gamma]$ is a C^*-norm, and the closure of $\mathbb{C}[\Gamma]$ with respect to this norm is a C^*-algebra, denoted $C^*(\Gamma)$.*

(2) *When Γ is abelian, we have an isomorphism $C^*(\Gamma) \simeq C(G)$, where $G = \widehat{\Gamma}$ is its Pontrjagin dual, formed by the characters $\chi : \Gamma \to \mathbb{T}$.*

Proof. All this is very standard, the idea being as follows:

(1) In order to prove the result, we must find a $*$-algebra embedding $\mathbb{C}[\Gamma] \subset B(H)$, with H being a Hilbert space. For this purpose, consider the space $H = l^2(\Gamma)$, having $\{h\}_{h\in\Gamma}$ as orthonormal basis. Our claim is that we have an embedding, as follows:

$$\pi : \mathbb{C}[\Gamma] \subset B(H) \quad , \quad \pi(g)(h) = gh$$

Indeed, since $\pi(g)$ maps the basis $\{h\}_{h\in\Gamma}$ into itself, this operator is well-defined, bounded, and is an isometry. It is also clear from the formula $\pi(g)(h) = gh$ that $g \to \pi(g)$ is a morphism of algebras, and since this morphism maps the unitaries $g \in \Gamma$ into isometries, this is a morphism of $*$-algebras. Finally, the faithfulness of π is clear.

(2) Since Γ is abelian, the corresponding group algebra $A = C^*(\Gamma)$ is commutative. Thus, we can apply the Gelfand theorem, and we obtain $A = C(X)$, with:

$$X = Spec(A)$$

But the spectrum $X = Spec(A)$, consisting of the characters $\chi : C^*(\Gamma) \to \mathbb{C}$, can be identified with the Pontrjagin dual $G = \widehat{\Gamma}$, and this gives the result. □

The above result suggests the following definition:

Definition 16.9. *Given a discrete group Γ, the compact quantum space G given by*

$$C(G) = C^*(\Gamma)$$

is called abstract dual of Γ, and is denoted $G = \widehat{\Gamma}$.

Good news, this definition is exactly what we need, in order to understand the meaning of Definitions 16.2 and 16.3. To be more precise, we have the following result:

Proposition 16.10. *The basic tori are all group duals, as follows,*

$$\begin{tikzcd} T_N^+ \arrow[r] & \mathbb{T}_N^+ & & \widehat{L_N} \arrow[r] & \widehat{F_N} \\ & & = & & \\ T_N \arrow[r] \arrow[uu] & \mathbb{T}_N \arrow[uu] & & \mathbb{Z}_2^N \arrow[r] \arrow[uu] & \mathbb{T}^N \arrow[uu] \end{tikzcd}$$

where $F_N = \mathbb{Z}^{*N}$ is the free group on N generators, and $L_N = \mathbb{Z}_2^{*N}$ is its real version.

Proof. The basic tori appear indeed as group duals, and together with the Fourier transform identifications from Theorem 16.8 (2), this gives the result. □

Moving ahead, now that we have our formalism, we can start developing free geometry. As a first objective, we would like to better understand the relation between the classical and free tori. In order to discuss this, let us introduce the following notion:

Definition 16.11. *Given a compact quantum space X, its classical version is the usual compact space $X_{class} \subset X$ obtained by dividing $C(X)$ by its commutator ideal:*

$$C(X_{class}) = C(X)/I \quad , \quad I =< [a,b] >$$

In this situation, we also say that X appears as a "liberation" of X.

In other words, the space X_{class} appears as the Gelfand spectrum of the commutative C^*-algebra $C(X)/I$. Observe in particular that X_{class} is indeed a classical space.

In relation now with our tori, we have the following result:

Theorem 16.12. *We have inclusions between the various tori, as follows,*

$$\begin{tikzcd} T_N^+ \arrow[r] & \mathbb{T}_N^+ \\ T_N \arrow[r] \arrow[u] & \mathbb{T}_N \arrow[u] \end{tikzcd}$$

and the free tori on top appear as liberations of the tori on the bottom.

Proof. This is indeed clear from definitions, because commutativity of a group algebra means precisely that the group in question is abelian. □

16b. Free spheres

In order to extend now the free geometries that we have, real and complex, let us begin with the spheres. We have the following notions:

DEFINITION 16.13. *We have free real and complex spheres, defined via*

$$C(S^{N-1}_{\mathbb{R},+})=C^*\left(x_1,\ldots,x_N\Big|x_i=x_i^*,\sum_ix_i^2=1\right)$$

$$C(S^{N-1}_{\mathbb{C},+})=C^*\left(x_1,\ldots,x_N\Big|\sum_ix_ix_i^*=\sum_ix_i^*x_i=1\right)$$

where the symbol C^* stands for universal enveloping C^*-algebra.

Here the fact that these algebras are indeed well-defined comes from the following estimate, which shows that the biggest C^*-norms on these $*$-algebras are bounded:

$$||x_i||^2=||x_ix_i^*||\leq\left|\left|\sum_ix_ix_i^*\right|\right|=1$$

As a first result now, regarding the above free spheres, we have:

THEOREM 16.14. *We have embeddings of compact quantum spaces, as follows,*

$$\begin{tikzcd}
S^{N-1}_{\mathbb{R},+} \arrow[r] & S^{N-1}_{\mathbb{C},+} \\
S^{N-1}_{\mathbb{R}} \arrow[r] \arrow[u] & S^{N-1}_{\mathbb{C}} \arrow[u]
\end{tikzcd}$$

and the spaces on top appear as liberations of the spaces on the bottom.

PROOF. The first assertion, regarding the inclusions, comes from the fact that at the level of the associated C^*-algebras, we have surjective maps, as follows:

$$\begin{tikzcd}
C(S^{N-1}_{\mathbb{R},+}) \arrow[d] & C(S^{N-1}_{\mathbb{C},+}) \arrow[l] \arrow[d] \\
C(S^{N-1}_{\mathbb{R}}) & C(S^{N-1}_{\mathbb{C}}) \arrow[l]
\end{tikzcd}$$

For the second assertion, we must establish the following isomorphisms, where the symbol C^*_{comm} stands for "universal commutative C^*-algebra generated by":

$$C(S^{N-1}_\mathbb{R})=C^*_{comm}\left(x_1,\ldots,x_N\Big|x_i=x_i^*,\sum_i x_i^2=1\right)$$

$$C(S^{N-1}_\mathbb{C})=C^*_{comm}\left(x_1,\ldots,x_N\Big|\sum_i x_ix_i^*=\sum_i x_i^*x_i=1\right)$$

It is enough to establish the second isomorphism. So, consider the second universal commutative C^*-algebra A constructed above. Since the standard coordinates on $S^{N-1}_\mathbb{C}$ satisfy the defining relations for A, we have a quotient map of as follows:

$$A\to C(S^{N-1}_\mathbb{C})$$

Conversely, let us write $A=C(S)$, by using the Gelfand theorem. Then $x_1,\ldots,x_N$ become in this way true coordinates, providing us with an embedding as follows:

$$S\subset\mathbb{C}^N$$

Also, the quadratic relations become $\sum_i|x_i|^2=1$, so we have $S\subset S^{N-1}_\mathbb{C}$. Thus, we have a quotient map $C(S^{N-1}_\mathbb{C})\to A$, as desired, and this gives all the results. □

By using the free spheres constructed above, we can now formulate:

Definition 16.15. *A real algebraic manifold $X\subset S^{N-1}_{\mathbb{C},+}$ is a closed quantum subspace defined, at the level of the corresponding C^*-algebra, by a formula of type*

$$C(X)=C(S^{N-1}_{\mathbb{C},+})\Big/\Big\langle f_i(x_1,\ldots,x_N)=0\Big\rangle$$

for certain family of noncommutative polynomials, as follows:

$$f_i\in\mathbb{C}<x_1,\ldots,x_N>$$

We denote by $\mathcal{C}(X)$ the $*$-subalgebra of $C(X)$ generated by the coordinates $x_1,\ldots,x_N$.

As a basic example here, we have the free real sphere $S^{N-1}_{\mathbb{R},+}$. The classical spheres $S^{N-1}_\mathbb{C},S^{N-1}_\mathbb{R}$, and their real submanifolds, are covered as well by this formalism. At the level of the general theory, we have the following version of the Gelfand theorem:

Theorem 16.16. *If $X\subset S^{N-1}_{\mathbb{C},+}$ is an algebraic manifold, as above, we have*

$$X_{class}=\left\{x\in S^{N-1}_\mathbb{C}\Big|f_i(x_1,\ldots,x_N)=0\right\}$$

and X appears as a liberation of X_{class}.

Proof. This is something that we already met, in the context of the free spheres. In general, the proof is similar, by using the Gelfand theorem. Indeed, if we denote by X_{class}' the manifold constructed in the statement, then we have a quotient map of C^*-algebras as follows, mapping standard coordinates to standard coordinates:

$$C(X_{class})\to C(X_{class}')$$

Conversely now, from $X\subset S^{N-1}_{\mathbb{C},+}$ we obtain $X_{class}\subset S^{N-1}_{\mathbb{C}}$. Now since the relations defining X_{class}' are satisfied by X_{class}, we obtain an inclusion $X_{class}\subset X_{class}'$. Thus, at the level of algebras of continuous functions, we have a quotient map of C^*-algebras as follows, mapping standard coordinates to standard coordinates:

$$C(X_{class}')\to C(X_{class})$$

Thus, we have constructed a pair of inverse morphisms, and we are done. □

Finally, once again at the level of the general theory, we have:

Definition 16.17. *We agree to identify two real algebraic submanifolds $X,Y\subset S^{N-1}_{\mathbb{C},+}$ when we have a $*$-algebra isomorphism between $*$-algebras of coordinates*

$$f:\mathcal{C}(Y)\to\mathcal{C}(X)$$

mapping standard coordinates to standard coordinates.

We will see later the reasons for making this convention, coming from amenability. Now back to the tori, as constructed before, we can see that these are examples of algebraic manifolds, in the sense of Definition 16.15. In fact, we have the following result:

Theorem 16.18. *The four main quantum spheres produce the main quantum tori*

$$\begin{tikzcd}
S^{N-1}_{\mathbb{R},+} \arrow[r] & S^{N-1}_{\mathbb{C},+} & & T_N^+ \arrow[r] & \mathbb{T}_N^+ \\
 & & \rightarrow & & \\
S^{N-1}_{\mathbb{R}} \arrow[r] \arrow[uu] & S^{N-1}_{\mathbb{C}} \arrow[uu] & & T_N \arrow[r] \arrow[uu] & \mathbb{T}_N \arrow[uu]
\end{tikzcd}$$

via the formula $T=S\cap\mathbb{T}_N^+$, with the intersection being taken inside $S^{N-1}_{\mathbb{C},+}$.

Proof. This comes from the above results, the situation being as follows:

(1) Free complex case. Here the formula in the statement reads:

$$\mathbb{T}_N^+=S^{N-1}_{\mathbb{C},+}\cap\mathbb{T}_N^+$$

But this is something trivial, because we have $\mathbb{T}_N^+\subset S^{N-1}_{\mathbb{C},+}$.

(2) Free real case. Here the formula in the statement reads:

$$T_N^+=S^{N-1}_{\mathbb{R},+}\cap\mathbb{T}_N^+$$

But this is clear as well, the real version of $\mathbb{T}_N^+$ being T_N^+.

(3) Classical complex case. Here the formula in the statement reads:

$$\mathbb{T}_N = S^{N-1}_\mathbb{C} \cap \mathbb{T}_N^+$$

But this is clear as well, the classical version of $\mathbb{T}_N^+$ being $\mathbb{T}_N$.

(4) Classical real case. Here the formula in the statement reads:

$$T_N = S^{N-1}_\mathbb{R} \cap \mathbb{T}_N^+$$

But this follows by intersecting the formulae from the proof of (2) and (3). □

In order to better understand the structure of the free spheres $S^{N-1}_{\mathbb{R},+}, S^{N-1}_{\mathbb{C},+}$, we need to talk about free rotations. Following Woronowicz, let us start with:

Definition 16.19. *A Woronowicz algebra is a C^*-algebra A, given with a unitary matrix $u \in M_N(A)$ whose coefficients generate A, such that the formulae*

$$\Delta(u_{ij}) = \sum_k u_{ik} \otimes u_{kj} \quad , \quad \varepsilon(u_{ij}) = \delta_{ij} \quad , \quad S(u_{ij}) = u_{ji}^*$$

define morphisms of C^*-algebras $\Delta : A \to A \otimes A$, $\varepsilon : A \to \mathbb{C}$, $S : A \to A^{opp}$.

We say that A is cocommutative when $\Sigma\Delta = \Delta$, where $\Sigma(a \otimes b) = b \otimes a$ is the flip. We have the following result, which justifies the terminology and axioms:

Theorem 16.20. *The following are Woronowicz algebras:*

(1) *$C(G)$, with $G \subset U_N$ compact Lie group. Here the structural maps are:*

$$\Delta(\varphi) = (g,h) \to \varphi(gh) \quad , \quad \varepsilon(\varphi) = \varphi(1) \quad , \quad S(\varphi) = g \to \varphi(g^{-1})$$

(2) *$C^*(\Gamma)$, with $F_N \to \Gamma$ finitely generated group. Here the structural maps are:*

$$\Delta(g) = g \otimes g \quad , \quad \varepsilon(g) = 1 \quad , \quad S(g) = g^{-1}$$

Moreover, we obtain in this way all the commutative/cocommutative algebras.

Proof. This is something very standard, the idea being as follows:

(1) Given $G \subset U_N$, we can set $A = C(G)$, which is a Woronowicz algebra, together with the matrix $u = (u_{ij})$ formed by coordinates of G, given by:

$$g = \begin{pmatrix} u_{11}(g) & \dots & u_{1N}(g) \\ \vdots & & \vdots \\ u_{N1}(g) & \dots & u_{NN}(g) \end{pmatrix}$$

Conversely, if (A, u) is a commutative Woronowicz algebra, by using the Gelfand theorem we can write $A = C(X)$, with X being a certain compact space. The coordinates u_{ij} give then an embedding $X \subset M_N(\mathbb{C})$, and since the matrix $u = (u_{ij})$ is unitary we actually obtain an embedding $X \subset U_N$, and finally by using the maps Δ, ε, S we conclude that our compact subspace $X \subset U_N$ is in fact a compact Lie group, as desired.

(2) Consider a finitely generated group $F_N \to \Gamma$. We can set $A = C^*(\Gamma)$, which is by definition the completion of the complex group algebra $\mathbb{C}[\Gamma]$, with involution given by $g^* = g^{-1}$, for any $g \in \Gamma$, with respect to the biggest C^*-norm, and we obtain a Woronowicz algebra, together with the diagonal matrix formed by the generators of Γ:

$$u = \begin{pmatrix} g_1 && 0 \\ & \ddots & \\ 0 && g_N \end{pmatrix}$$

Conversely, if (A, u) is a cocommutative Woronowicz algebra, the Peter-Weyl theory of Woronowicz, to be explained below, shows that the irreducible corepresentations of A are all 1-dimensional, and form a group Γ, and so we have $A = C^*(\Gamma)$, as desired. □

The above result makes it quite clear that what we have in Definition 16.19 is some sort of joint definition for the compact and discrete quantum groups. In order to further comment on this, let us go back to the formulae in Definition 16.19, namely:

$$\Delta(u_{ij}) = \sum_k u_{ik} \otimes u_{kj} \quad , \quad \varepsilon(u_{ij}) = \delta_{ij} \quad , \quad S(u_{ij}) = u_{ji}^*$$

The morphisms Δ, ε, S are called comultiplication, counit and antipode, and they have the following properties, which are something very familiar in abstract algebra:

Theorem 16.21. *Let (A, u) be a Woronowicz algebra.*

(1) *Δ, ε satisfy the usual axioms for a comultiplication and a counit, namely:*

$$\begin{aligned} (\Delta \otimes id)\Delta &= (id \otimes \Delta)\Delta \\ (\varepsilon \otimes id)\Delta &= (id \otimes \varepsilon)\Delta = id \end{aligned}$$

(2) *S satisfies the antipode axiom, on the $*$-subalgebra generated by entries of u:*

$$m(S \otimes id)\Delta = m(id \otimes S)\Delta = \varepsilon(.)1$$

(3) *In addition, the square of the antipode is the identity, $S^2 = id$.*

Proof. The two comultiplication axioms can be established as follows:

$$\begin{aligned} (\Delta \otimes id)\Delta(u_{ij}) &= (id \otimes \Delta)\Delta(u_{ij}) = \sum_{kl} u_{ik} \otimes u_{kl} \otimes u_{lj} \\ (\varepsilon \otimes id)\Delta(u_{ij}) &= (id \otimes \varepsilon)\Delta(u_{ij}) = u_{ij} \end{aligned}$$

As for the two antipode formulae, their verification is similar. □

Summarizing, the Woronowicz algebras appear to have nice properties. In view of Theorem 16.20 and Theorem 16.21, we can formulate the following definition:

Definition 16.22. *Given a Woronowicz algebra A, we formally write*

$$A=C(G)=C^*(\Gamma)$$

and call G compact quantum group, and Γ discrete quantum group.

In relation with this, there are actually some analytic subtleties, coming from amenability, so our objects must be divided by a certain equivalence relation, for everything to work fine. To be more precise, we agree to write $(A,u)=(B,v)$ when there is a $*$-algebra isomorphism as follows, mapping standard coordinates to standard coordinates:

$$<u_{ij}>\simeq<v_{ij}>\quad,\quad u_{ij}\to v_{ij}$$

Moving ahead now, let us call now corepresentation of A any unitary matrix $v\in M_n(A)$ satisfying the same conditions as those satisfied by u, namely:

$$\Delta(v_{ij})=\sum_k v_{ik}\otimes v_{kj}\quad,\quad\varepsilon(v_{ij})=\delta_{ij}\quad,\quad S(v_{ij})=v_{ji}^*$$

These corepresentations can be thought of as corresponding representations of the underlying compact quantum group G. Following Woronowicz, we have:

Theorem 16.23. *Any Woronowicz algebra has a unique Haar integration functional,*

$$\left(\int_G\otimes id\right)\Delta=\left(id\otimes\int_G\right)\Delta=\int_G(.)1$$

which can be constructed by starting with any faithful positive form $\varphi\in A^*$, and setting

$$\int_G=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\varphi^{*k}$$

where $\phi*\psi=(\phi\otimes\psi)\Delta$. Moreover, for any corepresentation $v\in M_n(\mathbb C)\otimes A$ we have

$$\left(id\otimes\int_G\right)v=P$$

where P is the orthogonal projection onto $Fix(v)=\{\xi\in\mathbb C^n|v\xi=\xi\}$.

Proof. This can be done in 3 steps, as follows:

(1) Given $\varphi\in A^*$, our claim is that the following limit converges, for any $a\in A$:

$$\int_\varphi a=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\varphi^{*k}(a)$$

Indeed, by linearity we can assume that a is the coefficient of corepresentation, $a=(\tau\otimes id)v$. But in this case, an elementary computation shows that we have the following formula, where P_φ is the orthogonal projection onto the 1-eigenspace of $(id\otimes\varphi)v$:

$$\left(id\otimes\int_\varphi\right)v=P_\varphi$$

(2) Since $v\xi=\xi$ implies $[(id\otimes\varphi)v]\xi=\xi$, we have $P_\varphi\geq P$, where P is the orthogonal projection onto the space $Fix(v)=\{\xi\in\mathbb C^n|v\xi=\xi\}$. The point now is that when $\varphi\in A^*$ is faithful, by using a positivity trick, one can prove that we have $P_\varphi=P$. Thus our linear form $\int_\varphi$ is independent of φ, and is given on coefficients $a=(\tau\otimes id)v$ by:

$$\left(id\otimes\int_\varphi\right)v=P$$

(3) With the above formula in hand, the left and right invariance of $\int_G=\int_\varphi$ is clear on coefficients, and so in general, and this gives all the assertions. □

Consider the dense $*$-subalgebra $\mathcal A\subset A$ generated by the coefficients of u, and endow it with the scalar product $<a,b>=\int_G ab^*$. Still following Woronowicz, we have:

Theorem 16.24. *We have the following Peter-Weyl type results:*

1. *Any corepresentation decomposes as a sum of irreducible corepresentations.*
2. *Each irreducible corepresentation appears inside a certain $u^{\otimes k}$.*
3. *$\mathcal A=\bigoplus_{v\in Irr(A)}M_{\dim(v)}(\mathbb C)$, the summands being pairwise orthogonal.*
4. *The characters of irreducible corepresentations form an orthonormal system.*

Proof. All these results are very standard, the idea being as follows:

(1) Given $v\in M_n(A)$, its intertwiner algebra $End(v)=\{T\in M_n(\mathbb C)|Tv=vT\}$ is a finite dimensional C^*-algebra, and so decomposes as $End(v)=M_{n_1}(\mathbb C)\oplus\ldots\oplus M_{n_r}(\mathbb C)$. But this gives a decomposition of type $v=v_1+\ldots+v_r$, as desired.

(2) Consider indeed the Peter-Weyl corepresentations, $u^{\otimes k}$ with k colored integer, defined by $u^{\otimes\emptyset}=1$, $u^{\otimes\circ}=u$, $u^{\otimes\bullet}=\bar u$ and multiplicativity. The coefficients of these corepresentations span the dense algebra $\mathcal A$, and by using (1), this gives the result.

(3) Here the direct sum decomposition, which is technically a $*$-coalgebra isomorphism, follows from (2). As for the second assertion, this follows from the fact that $(id\otimes\int_G)v$ is the orthogonal projection P_v onto the space $Fix(v)$, for any corepresentation v.

(4) Let us define indeed the character of $v\in M_n(A)$ to be the matrix trace, $\chi_v=Tr(v)$. Since this character is a coefficient of v, the orthogonality assertion follows from (3). As for the norm 1 claim, this follows once again from $(id\otimes\int_G)v=P_v$. □

Good news, we can now talk about free rotations. Following Wang, we have:

Theorem 16.25. *The following universal algebras are Woronowicz algebras,*

$$C(O_N^+)=C^*\left((u_{ij})_{i,j=1,\ldots,N}\Big|u=\bar u,u^t=u^{-1}\right)$$

$$C(U_N^+)=C^*\left((u_{ij})_{i,j=1,\ldots,N}\Big|u^*=u^{-1},u^t=\bar u^{-1}\right)$$

so the underlying spaces O_N^+,U_N^+ are compact quantum groups.

PROOF. This follows from the elementary fact that if a matrix $u = (u_{ij})$ is orthogonal or biunitary, then so must be the following matrices:

$$u_{ij}^\Delta = \sum_k u_{ik} \otimes u_{kj} \quad , \quad u_{ij}^\varepsilon = \delta_{ij} \quad , \quad u_{ij}^S = u_{ji}^*$$

Thus, we can indeed define morphisms Δ, ε, S as in Definition 16.19, by using the universal properties of $C(O_N^+)$, $C(U_N^+)$, and this gives the result. □

Let us discuss now the correspondence $U \to S$. In the classical case the situation is very simple, because the sphere $S = S^{N-1}$ appears by rotating the point $x = (1, 0, \ldots, 0)$ by the isometries in $U = U_N$. Moreover, the stabilizer of this action is the subgroup $U_{N-1} \subset U_N$ acting on the last $N-1$ coordinates, and so the sphere $S = S^{N-1}$ appears from the corresponding rotation group $U = U_N$ as an homogeneous space, as follows:

$$S^{N-1} = U_N / U_{N-1}$$

In functional analytic terms, all this becomes even simpler, the correspondence $U \to S$ being obtained, at the level of algebras of functions, as follows:

$$C(S^{N-1}) \subset C(U_N) \quad , \quad x_i \to u_{1i}$$

In general now, the straightforward homogeneous space interpretation of S as above fails. However, we can have some theory going by using the functional analytic viewpoint, with an embedding $x_i \to u_{1i}$ as above. Let us start with the following result:

PROPOSITION 16.26. *For the basic spheres, we have a diagram as follows,*

$$\begin{tikzcd}
C(S) \arrow[r, "\Phi"] \arrow[d, "\alpha"] & C(S) \otimes C(U) \arrow[d, "\alpha\otimes id"] \\
C(U) \arrow[r, "\Delta"] & C(U) \otimes C(U)
\end{tikzcd}$$

where on top $\Phi(x_i) = \sum_j x_j \otimes u_{ji}$, and on the left $\alpha(x_i) = u_{1i}$.

PROOF. The diagram in the statement commutes indeed on the standard coordinates, the corresponding arrows being as follows, on these coordinates:

$$\begin{tikzcd}
x_i \arrow[r] \arrow[d] & \sum_j x_j \otimes u_{ji} \arrow[d] \\
u_{1i} \arrow[r] & \sum_j u_{1j} \otimes u_{ji}
\end{tikzcd}$$

Thus by linearity and multiplicativity, the whole the diagram commutes. □

As a consequence of the above result, we can now formulate:

Proposition 16.27. *We have a quotient map and an inclusion as follows,*

$$U \to S_U \subset S$$

with S_U being the first row space of U, given by

$$C(S_U) =< u_{1i} >\subset C(U)$$

at the level of the corresponding algebras of functions.

Proof. At the algebra level, we have an inclusion and a quotient map as follows:

$$C(S) \to C(S_U) \subset C(U)$$

Thus, we obtain the result, by transposing. □

In order to advance, we will use the uniform integration over S, which can be introduced, in analogy with what happens in the classical case, in the following way:

Definition 16.28. *We endow each of the algebras $C(S)$ with its integration functional*

$$\int_S : C(S) \to C(U) \to \mathbb{C}$$

obtained by composing the morphism $x_i \to u_{1i}$ with the Haar integration of $C(U)$.

With this in hand, we can now integrate over the spheres S, as follows:

Theorem 16.29. *The integration over the basic spheres is given by*

$$\int_S x_{i_1}^{e_1} \ldots x_{i_k}^{e_k} = \sum_\pi \sum_{\sigma \leq \ker i} W_{kN}(\pi, \sigma)$$

with $\pi, \sigma \in D(k)$, where $W_{kN} = G_{kN}^{-1}$ is the inverse of $G_{kN}(\pi, \sigma) = N^{|\pi \vee \sigma|}$.

Proof. According to our conventions, the integration over S is a particular case of the integration over U, via $x_i = u_{1i}$. By using the Weingarten formula, we obtain:

$$\begin{aligned}
\int_S x_{i_1}^{e_1} \ldots x_{i_k}^{e_k} &= \int_U u_{1i_1}^{e_1} \ldots u_{1i_k}^{e_k} \\
&= \sum_{\pi, \sigma \in D(k)} \delta_\pi(1) \delta_\sigma(i) W_{kN}(\pi, \sigma) \\
&= \sum_{\pi, \sigma \in D(k)} \delta_\sigma(i) W_{kN}(\pi, \sigma)
\end{aligned}$$

Thus, we are led to the formula in the statement. □

16c. Projective spaces

Getting back now to our projective business, our starting point will be the following functional analytic description of the real and complex projective spaces $P^{N-1}_\mathbb{R},P^{N-1}_\mathbb{C}$:

Theorem 16.30. *We have presentation results as follows,*

$$\begin{aligned}C(P^{N-1}_\mathbb{R})&=&C^*_{comm}\left((p_{ij})_{i,j=1,\ldots,N}\Big|p=\bar{p}=p^t=p^2,Tr(p)=1\right)\\C(P^{N-1}_\mathbb{C})&=&C^*_{comm}\left((p_{ij})_{i,j=1,\ldots,N}\Big|p=p^*=p^2,Tr(p)=1\right)\end{aligned}$$

for the algebras of continuous functions on the real and complex projective spaces.

Proof. We use the fact that the projective spaces $P^{N-1}_\mathbb{R},P^{N-1}_\mathbb{C}$ can be respectively identified with the spaces of rank one projections in $M_N(\mathbb{R}),M_N(\mathbb{C})$. With this picture in mind, it is clear that we have arrows $\leftarrow$. In order to construct now arrows $\to$, consider the universal algebras on the right, A_R,A_C. These algebras being both commutative, by the Gelfand theorem we can write, with X_R,X_C being certain compact spaces:

$$A_R=C(X_R)\quad,\quad A_C=C(X_C)$$

Now by using the coordinate functions p_{ij}, we conclude that X_R,X_C are certain spaces of rank one projections in $M_N(\mathbb{R}),M_N(\mathbb{C})$. In other words, we have embeddings:

$$X_R\subset P^{N-1}_\mathbb{R}\quad,\quad X_C\subset P^{N-1}_\mathbb{C}$$

By transposing we obtain arrows $\to$, as desired. □

The point now is that the above result suggests the following definition:

Definition 16.31. *Associated to any $N\in\mathbb{N}$ is the following universal algebra,*

$$C(P^{N-1}_+)=C^*\left((p_{ij})_{i,j=1,\ldots,N}\Big|p=p^*=p^2,Tr(p)=1\right)$$

whose abstract spectrum is called "free projective space".

Observe that, according to our presentation results for the real and complex projective spaces $P^{N-1}_\mathbb{R}$ and $P^{N-1}_\mathbb{C}$, we have embeddings of compact quantum spaces, as follows:

$$P^{N-1}_\mathbb{R}\subset P^{N-1}_\mathbb{C}\subset P^{N-1}_+$$

Let us first discuss the relation with the spheres. Given a closed subset $X\subset S^{N-1}_{\mathbb{R},+}$, its projective version is by definition the quotient space $X\to PX$ determined by the fact that $C(PX)\subset C(X)$ is the subalgebra generated by the following variables:

$$p_{ij}=x_ix_j$$

In order to discuss the relation with the spheres, let us start with:

Theorem 16.32. *The projective versions of the* 3 *real spheres are as follows,*

$$\begin{tikzcd}
S^{N-1}_{\mathbb{R}} \arrow[r] \arrow[d] & S^{N-1}_{\mathbb{R},*} \arrow[r] \arrow[d] & S^{N-1}_{\mathbb{R},+} \arrow[d] \\
P^{N-1}_{\mathbb{R}} \arrow[r] & P^{N-1}_{\mathbb{C}} \arrow[r] & P^{N-1}_{+}
\end{tikzcd}$$

modulo the standard equivalence relation for the quantum algebraic manifolds.

Proof. The assertion at left is true by definition. For the assertion at right, we have to prove that the variables $p_{ij} = z_i z_j$ over the free sphere $S^{N-1}_{\mathbb{R},+}$ satisfy the defining relations for $C(P^{N-1}_+)$, from Definition 16.31, namely:

$$p = p^* = p^2 \quad , \quad Tr(p) = 1$$

We first have the following computation:

$$(p^*)_{ij} = p^*_{ji} = (z_j z_i)^* = z_i z_j = p_{ij}$$

We have as well the following computation:

$$(p^2)_{ij} = \sum_k p_{ik} p_{kj} = \sum_k z_i z_k^2 z_j = z_i z_j = p_{ij}$$

Finally, we have as well the following computation:

$$Tr(p) = \sum_k p_{kk} = \sum_k z_k^2 = 1$$

Regarding now $PS^{N-1}_{\mathbb{R},*} = P^{N-1}_{\mathbb{C}}$, the inclusion "$\subset$" follows from $abcd = cbad = cbda$. In the other sense now, the point is that we have a matrix model, as follows:

$$\pi : C(S^{N-1}_{\mathbb{R},*}) \to M_2(C(S^{N-1}_{\mathbb{C}})) \quad , \quad x_i \to \begin{pmatrix} 0 & z_i \\ \bar{z}_i & 0 \end{pmatrix}$$

But this gives the missing inclusion "$\supset$", and we are done. □

16d. Threefold way

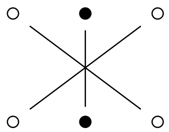
Looking back at the definition of the spheres that we have, and at the precise relations between the coordinates, we are led into the following notion:

Definition 16.33. *A monomial sphere is a subset $S \subset S^{N-1}_{\mathbb{C},+}$ obtained via relations*

$$x_{i_1}^{e_1} \ldots x_{i_k}^{e_k} = x_{i_{\sigma(1)}}^{f_1} \ldots x_{i_{\sigma(k)}}^{f_k} \quad , \quad \forall (i_1, \ldots, i_k) \in \{1, \ldots, N\}^k$$

with $\sigma \in S_k$ being certain permutations, and with $e_r, f_r \in \{1, *\}$ being certain exponents.

This definition is quite broad, and we have for instance as example the sphere $S^{N-1}_{\mathbb{C},\times}$ coming from the relations $ab^*c=cb^*a$, corresponding to the following diagram:



In view of these difficulties, we will restrict now the attention to the real case. Let us first recall that we have the following fundamental result, dealing with the real case:

Theorem 16.34. *There are exactly* 3 *real easy geometries, namely*

$$\mathbb{R}^N\subset\mathbb{R}^N_*\subset\mathbb{R}^N_+$$

coming from $P_2\supset P_2^*\supset NC_2$, whose associated spheres are

$$S^{N-1}_\mathbb{R}\subset S^{N-1}_{\mathbb{R},*}\subset S^{N-1}_{\mathbb{R},+}$$

and whose tori, unitary and reflection groups are given by similar formulae.

Proof. This is something that we know well, coming from the fact that $G=O_N^*$ is the unique intermediate easy quantum group $O_N\subset G\subset O_N^+$. □

Let us focus now on the spheres, and try to better understand their "easiness" property, with results getting beyond what has been done above, in the general easy context. That is, our objects of interest in what follows will be the 3 real spheres, namely:

$$S^{N-1}_\mathbb{R}\subset S^{N-1}_{\mathbb{R},*}\subset S^{N-1}_{\mathbb{R},+}$$

Our purpose in what follows we will be that of proving that these spheres are the only monomial ones. In order to best talk about monomiality, in the present real case, it is convenient to introduce the following group:

$$S_\infty=\bigcup_{k\geq0}S_k$$

To be more precise, this group appears by definition as an inductive limit, with the inclusions $S_k\subset S_{k+1}$ that we use being given by:

$$\sigma\in S_k\implies\sigma(k+1)=k+1$$

In terms of S_∞, the definition of the monomial spheres reformulates as follows:

Proposition 16.35. *The monomial spheres are the algebraic manifolds $S\subset S^{N-1}_{\mathbb{R},+}$ obtained via relations of type*

$$x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}},\ \forall(i_1,\ldots,i_k)\in\{1,\ldots,N\}^k$$

associated to certain elements $\sigma\in S_\infty$, where $k\in\mathbb{N}$ is such that $\sigma\in S_k$.

Proof. We must prove that the relations $x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}$ are left unchanged when replacing $k\to k+1$. But this follows from $\sum_i x_i^2=1$, because:

$$\begin{aligned}
&x_{i_1}\ldots x_{i_k}x_{i_{k+1}}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}x_{i_{k+1}}\\
\implies\ &x_{i_1}\ldots x_{i_k}x_{i_{k+1}}^2=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}x_{i_{k+1}}^2\\
\implies\ &\sum_{i_{k+1}}x_{i_1}\ldots x_{i_k}x_{i_{k+1}}^2=\sum_{i_{k+1}}x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}x_{i_{k+1}}^2\\
\implies\ &x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}
\end{aligned}$$

Thus we can indeed "simplify at right", and this gives the result. □

As already mentioned, our goal in what follows will be that of proving that the 3 main spheres are the only monomial ones. In order to prove this result, we will use group theory methods. We call a subgroup $G\subset S_\infty$ filtered when it is stable under concatenation, in the sense that when writing $G=(G_k)$ with $G_k\subset S_k$, we have:

$$\sigma\in G_k,\pi\in G_l\implies\sigma\pi\in G_{k+l}$$

With this convention, we have the following result:

Theorem 16.36. *The monomial spheres are the subsets $S_G\subset S^{N-1}_{\mathbb R,+}$ given by*

$$C(S_G)=C(S^{N-1}_{\mathbb R,+})\Big/\Big\langle x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}},\forall(i_1,\ldots,i_k)\in\{1,\ldots,N\}^k,\forall\sigma\in G_k\Big\rangle$$

where $G=(G_k)$ is a filtered subgroup of $S_\infty=(S_k)$.

Proof. We know from Proposition 16.35 that the construction in the statement produces a monomial sphere. Conversely, given a monomial sphere $S\subset S^{N-1}_{\mathbb R,+}$, let us set:

$$G_k=\left\{\sigma\in S_k\Big|x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}},\forall(i_1,\ldots,i_k)\in\{1,\ldots,N\}^k\right\}$$

With $G=(G_k)$ we have then $S=S_G$. Thus, it remains to prove that G is a filtered group. But since the relations $x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}$ can be composed and reversed, each G_k follows to be stable under composition and inversion, and is therefore a group. Also, since the relations $x_{i_1}\ldots x_{i_k}=x_{i_{\sigma(1)}}\ldots x_{i_{\sigma(k)}}$ can be concatenated as well, our group $G=(G_k)$ is stable under concatenation, and we are done. □

At the level of examples, according to our definitions, the simplest filtered groups, namely $\{1\}\subset S_\infty$, produce the simplest real spheres, namely:

$$S^{N-1}_{\mathbb R,+}\supset S^{N-1}_{\mathbb R}$$

In order to discuss now the half-classical case, we need to introduce and study a certain privileged intermediate filtered group $\{1\}\subset S_\infty^*\subset S_\infty$, which will eventually produce the intermediate sphere $S^{N-1}_{\mathbb R,+}\supset S^{N-1}_{\mathbb R,*}\supset S^{N-1}_{\mathbb R}$. This can be done as follows:

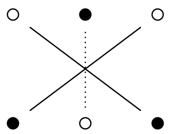
Proposition 16.37. *Let $S_\infty^* \subset S_\infty$ be the set of permutations having the property that when labelling cyclically the legs as follows*

$$\bullet\circ\bullet\circ\ldots$$

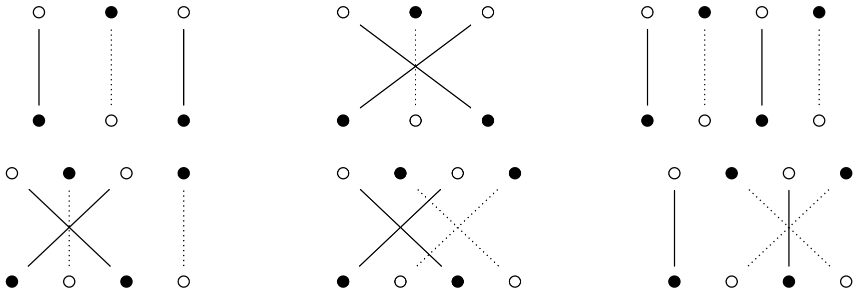
each string joins a black leg to a white leg.

1. *S_∞^* is a filtered subgroup of S_∞, generated by the half-classical crossing.*
2. *We have $S_{2k}^* \simeq S_k \times S_k$, and $S_{2k+1}^* \simeq S_k \times S_{k+1}$, for any $k \in \mathbb{N}$.*

Proof. The fact that S_∞^* is indeed a subgroup of S_∞, which is filtered, is clear. Observe now that the half-classical crossing has the "black-to-white" joining property:



Thus this crossing belongs to S_3^*, and it is routine to check that the filtered subgroup of S_∞ generated by it is the whole S_∞^*. Regarding now the last assertion, observe first that the filtered subgroups S_3^*, S_4^* consist of the following permutations:



Thus we have $S_3^* = S_1 \times S_2$ and $S_4^* = S_2 \times S_2$, with the first component coming from dotted permutations, and with the second component coming from the solid line permutations. The same argument works in general, and gives the last assertion. □

Now back to the main 3 real spheres, the result is as follows:

Proposition 16.38. *The basic monomial real spheres, namely*

$$S^{N-1}_{\mathbb{R}} \subset S^{N-1}_{\mathbb{R},*} \subset S^{N-1}_{\mathbb{R},+}$$

come respectively from the filtered groups $S_\infty \supset S_\infty^* \supset \{1\}$.

Proof. This is clear by definition in the classical and in the free cases. In the half-liberated case, the result follows from Proposition 16.37 (1). □

Now back to the general case, with the idea in mind of proving the uniqueness of the above spheres, consider a monomial sphere $S_G \subset S^{N-1}_{\mathbb{R},+}$, with the filtered group $G \subset S_\infty$ taken to be maximal, as in the proof of Theorem 16.36. We have the following result:

Proposition 16.39. *The filtered group $G\subset S_\infty$ associated to a monomial sphere $S\subset S^{N-1}_{\mathbb R,+}$ is stable under the following operations, on the corresponding diagrams:*

(1) *Removing outer strings.*
(2) *Removing neighboring strings.*

Proof. Both these results follow by using the quadratic condition:

(1) Regarding the outer strings, by summing over a, we have:

$$\begin{aligned}Xa=Ya&\implies Xa^2=Ya^2\\&\implies X=Y\end{aligned}$$

We have as well the following computation:

$$\begin{aligned}aX=aY&\implies a^2X=a^2Y\\&\implies X=Y\end{aligned}$$

(2) Regarding the neighboring strings, once again by summing over a, we have:

$$\begin{aligned}XabY=ZabT&\implies Xa^2Y=Za^2T\\&\implies XY=ZT\end{aligned}$$

We have as well the following computation:

$$\begin{aligned}XabY=ZbaT&\implies Xa^2Y=Za^2T\\&\implies XY=ZT\end{aligned}$$

Thus $G=(G_k)$ has both the properties in the statement. □

We can now state and prove a main result, as follows:

Theorem 16.40. *There is only one intermediate monomial sphere*

$$S^{N-1}_{\mathbb R}\subset S\subset S^{N-1}_{\mathbb R,+}$$

namely the half-classical real sphere $S^{N-1}_{\mathbb R,*}$.

Proof. We will prove that the only filtered groups $G\subset S_\infty$ satisfying the conditions in Proposition 16.39 are those correspoding to our 3 spheres, namely:

$$\{1\}\subset S_\infty^*\subset S_\infty$$

In order to do so, consider such a filtered group $G\subset S_\infty$. We assume this group to be non-trivial, $G\neq\{1\}$, and we want to prove that we have $G=S_\infty^*$ or $G=S_\infty$.

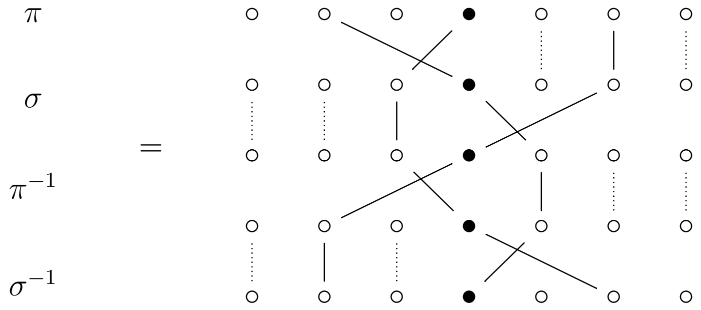
Step 1. Our first claim is that G contains a 3-cycle. Assume indeed that two permutations $\pi,\sigma\in S_\infty$ have support overlapping on exactly one point, say:

$$supp(\pi)\cap supp(\sigma)=\{i\}$$

The point is then that the commutator $\sigma^{-1}\pi^{-1}\sigma\pi$ is a 3-cycle, namely:

$$(i,\sigma^{-1}(i),\pi^{-1}(i))$$

Indeed the computation of the commutator goes as follows:

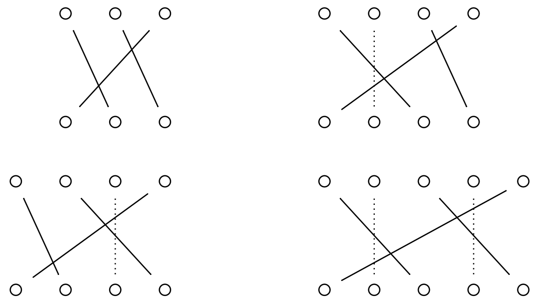


Now let us pick a non-trivial element $\tau\in G$. By removing outer strings at right and at left we obtain permutations $\tau'\in G_k,\tau''\in G_s$ having a non-trivial action on their right/left leg, and the trick applies, with:

$$\pi=\tau'\otimes id_{s-1}\quad,\quad\sigma=id_{k-1}\otimes\tau''$$

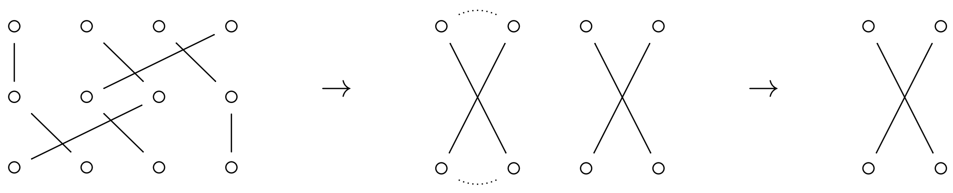
Thus, G contains a 3-cycle, as claimed.

Step 2. Our second claim is G must contain one of the following permutations:

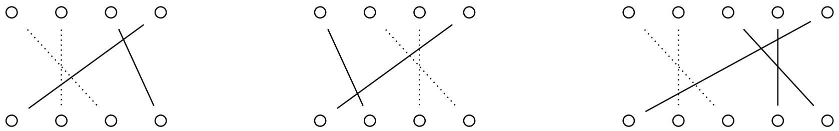


Indeed, consider the 3-cycle that we just constructed. By removing all outer strings, and then all pairs of adjacent vertical strings, we are left with these permutations.

Step 3. Our claim now is that we must have $S_\infty^*\subset G$. Indeed, let us pick one of the permutations that we just constructed, and apply to it our various diagrammatic rules. From the first permutation we can obtain the basic crossing, as follows:



Also, by removing a suitable ╳ shaped configuration, which is represented by dotted lines in the diagrams below, we can obtain the basic crossing from the second and third permutation, and the half-liberated crossing from the fourth permutation:



Thus, in all cases we have a basic or half-liberated crossing, and so, as desired:

$$S_\infty^* \subset G$$

Step 4. Our last claim, which will finish the proof, is that there is no proper intermediate subgroup as follows:

$$S_\infty^* \subset G \subset S_\infty$$

In order to prove this, observe that $S_\infty^* \subset S_\infty$ is the subgroup of parity-preserving permutations, in the sense that "i even $\implies \sigma(i)$ even".

Now let us pick an element $\sigma \in S_k - S_k^*$, with $k \in \mathbb{N}$. We must prove that the group $G =< S_\infty^*, \sigma >$ equals the whole S_∞. In order to do so, we use the fact that σ is not parity preserving. Thus, we can find i even such that $\sigma(i)$ is odd. In addition, up to passing to $\sigma|$, we can assume that $\sigma(k) = k$, and then, up to passing one more time to $\sigma|$, we can further assume that k is even. Since both i, k are even we have:

$$(i, k) \in S_k^*$$

We conclude that the following element belongs to G:

$$\sigma(i, k)\sigma^{-1} = (\sigma(i), k)$$

But, since $\sigma(i)$ is odd, by deleting an appropriate number of vertical strings, $(\sigma(i), k)$ reduces to the basic crossing $(1, 2)$. Thus $G = S_\infty$, and we are done. □

We have a similar result in the projective setting, as follows:

Theorem 16.41 (Threefold way). *The basic projective spaces, namely*

$$P_\mathbb{R}^{N-1} \subset P_\mathbb{C}^{N-1} \subset P_+^{N-1}$$

are the only monomial ones.

Proof. This follows indeed by using the same arguments as for the spheres. □

16e. Exercises

Congratulations for having read this book, and no exercises for this final chapter.

Bibliography

[1] V.I. Arnold, Ordinary differential equations, Springer (1973).

[2] V.I. Arnold, Mathematical methods of classical mechanics, Springer (1974).

[3] V.I. Arnold, Lectures on partial differential equations, Springer (1997).

[4] V.I. Arnold, Catastrophe theory, Springer (1974).

[5] V.I. Arnold and B.A. Khesin, Topological methods in hydrodynamics, Springer (1998).

[6] M.F. Atiyah, K-theory, CRC Press (1964).

[7] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).

[8] M.F. Atiyah and I.G. MacDonald, Introduction to commutative algebra, Addison-Wesley (1969).

[9] T. Banica, Calculus and applications (2024).

[10] T. Banica, Advanced linear algebra (2025).

[11] T. Banica, Geometry and topology (2025).

[12] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).

[13] N. Berline, E. Getzler and M. Vergne, Heat kernels and Dirac operators, Springer (2004).

[14] B. Blackadar, K-theory for operator algebras, Cambridge Univ. Press (1986).

[15] S.J. Blundell and K.M. Blundell, Concepts in thermal physics, Oxford Univ. Press (2006).

[16] S.M. Carroll, Spacetime and geometry, Cambridge Univ. Press (2004).

[17] A.R. Choudhuri, Astrophysics for physicists, Cambridge Univ. Press (2012).

[18] A. Connes, Noncommutative geometry, Academic Press (1994).

[19] A. Connes and M. Marcolli, Noncommutative geometry, quantum fields and motives, AMS (2008).

[20] W.N. Cottingham and D.A. Greenwood, An introduction to the standard model of particle physics, Cambridge Univ. Press (2012).

[21] P.A. Davidson, Introduction to magnetohydrodynamics, Cambridge Univ. Press (2001).

[22] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).

[23] M.P. do Carmo, Differential geometry of curves and surfaces, Dover (1976).

[24] M.P. do Carmo, Riemannian geometry, Birkhäuser (1992).

[25] S. Dodelson, Modern cosmology, Academic Press (2003).

[26] S.K. Donaldson, Riemann surfaces, Oxford Univ. Press (2004).

[27] R. Durrett, Probability: theory and examples, Cambridge Univ. Press (1990).

[28] A. Einstein, Relativity: the special and the general theory, Dover (1916).

[29] L.C. Evans, Partial differential equations, AMS (1998).

[30] W. Feller, An introduction to probability theory and its applications, Wiley (1950).

[31] E. Fermi, Thermodynamics, Dover (1937).

[32] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics, Caltech (1963).

[33] R.P. Feynman and A.R. Hibbs, Quantum mechanics and path integrals, Dover (1965).

[34] P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge Univ. Press (2009).

[35] A.P. French, Special relativity, Taylor and Francis (1968).

[36] W. Fulton, Algebraic topology, Springer (1995).

[37] W. Fulton and J. Harris, Representation theory, Springer (1991).

[38] C. Godsil and G. Royle, Algebraic graph theory, Springer (2001).

[39] H. Goldstein, C. Safko and J. Poole, Classical mechanics, Addison-Wesley (1980).

[40] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory, Cambridge Univ. Press (2012).

[41] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (2017).

[42] D.J. Griffiths and D.F. Schroeter, Introduction to quantum mechanics, Cambridge Univ. Press (2018).

[43] D.J. Griffiths, Introduction to elementary particles, Wiley (2020).

[44] P. Griffiths and J. Harris, Principles of algebraic geometry, Wiley (1994).

[45] A. Grothendieck and J. Dieudonné, Éléments de géométrie algébrique, IHES (1967).

[46] A. Grothendieck et al., Séminaire de géométrie algébrique, IHES (1972).

[47] G.H. Hardy and E.M. Wright, An introduction to the theory of numbers, Oxford Univ. Press (1938).

[48] J. Harris, Algebraic geometry, Springer (1992).

[49] R. Hartshorne, Algebraic geometry, Springer (1977).

[50] A. Hatcher, Algebraic topology, Cambridge Univ. Press (2002).

[51] H. Hofer and E. Zehnder, Symplectic invariants and Hamiltonian dynamics, Birkhäuser (1994).

[52] L. Hörmander, The analysis of linear partial differential operators, Springer (1983).

[53] R.A. Horn and C.R. Johnson, Matrix analysis, Cambridge Univ. Press (1985).

[54] K. Huang, Introduction to statistical physics, CRC Press (2001).

[55] J.E. Humphreys, Introduction to Lie algebras and representation theory, Springer (1972).

[56] J.E. Humphreys, Linear algebraic groups, Springer (1975).

[57] K. Ireland and M. Rosen, A classical introduction to modern number theory, Springer (1982).

[58] N. Jacobson, Basic algebra, Dover (1974).

[59] V.F.R. Jones, Index for subfactors, *Invent. Math.* **72** (1983), 1–25.

[60] V.F.R. Jones, A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* **12** (1985), 103–111.

[61] V.F.R. Jones, Hecke algebra representations of braid groups and link polynomials, *Ann. of Math.* **126** (1987), 335–388.

[62] V.F.R. Jones, On knot invariants related to some statistical mechanical models, *Pacific J. Math.* **137** (1989), 311–334.

[63] V.F.R. Jones, Planar algebras I (1999).

[64] M. Karoubi, K-theory: an introduction, Springer (1978).

[65] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).

[66] T. Lancaster and K.M. Blundell, Quantum field theory for the gifted amateur, Oxford Univ. Press (2014).

[67] L.D. Landau and E.M. Lifshitz, Course of theoretical physics, Pergamon Press (1960).

[68] S. Lang, Algebra, Addison-Wesley (1993).

[69] S. Lang, Abelian varieties, Dover (1959).

[70] P. Lax, Linear algebra and its applications, Wiley (2007).

[71] P. Lax, Functional analysis, Wiley (2002).

[72] J.M. Lee, Introduction to topological manifolds, Springer (2011).

[73] J.M. Lee, Introduction to smooth manifolds, Springer (2012).

[74] J.M. Lee, Introduction to Riemannian manifolds, Springer (2019).

[75] D. McDuff and D. Salamon, Introduction to symplectic topology, Oxford Univ. Press (2017).

[76] P. Petersen, Linear algebra, Springer (2012).

[77] P. Petersen, Riemannian geometry, Springer (2006).

[78] W. Rudin, Principles of mathematical analysis, McGraw-Hill (1964).

[79] W. Rudin, Real and complex analysis, McGraw-Hill (1966).

[80] W. Rudin, Fourier analysis on groups, Dover (1974).

[81] B. Ryden, Introduction to cosmology, Cambridge Univ. Press (2002).

[82] B. Ryden and B.M. Peterson, Foundations of astrophysics, Cambridge Univ. Press (2010).

[83] W. Schlag, A course in complex analysis and Riemann surfaces, AMS (2014).

[84] D.V. Schroeder, An introduction to thermal physics, Oxford Univ. Press (1999).

[85] J.P. Serre, A course in arithmetic, Springer (1973).

[86] J.P. Serre, Linear representations of finite groups, Springer (1977).

[87] I.R. Shafarevich, Basic algebraic geometry, Springer (1974).

[88] J.H. Silverman, The arithmetic of elliptic curves, Springer (1986).

[89] J.H. Silverman and J.T. Tate, Rational points on elliptic curves, Springer (2015).

[90] B. Singh, Basic commutative algebra, World Scientific (2011).

[91] C.H. Taubes, Differential geometry, Oxford Univ. Press (2011).

[92] J.R. Taylor, Classical mechanics, Univ. Science Books (2003).

[93] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press (1955).

[94] S. Weinberg, Foundations of modern physics, Cambridge Univ. Press (2011).

[95] S. Weinberg, Lectures on quantum mechanics, Cambridge Univ. Press (2012).

[96] S. Weinberg, Lectures on astrophysics, Cambridge Univ. Press (2019).

[97] H. Weyl, The theory of groups and quantum mechanics, Princeton Univ. Press (1931).

[98] H. Weyl, The classical groups: their invariants and representations, Princeton Univ. Press (1939).

[99] H. Weyl, Space, time, matter, Princeton Univ. Press (1918).

[100] B. Zwiebach, A first course in string theory, Cambridge Univ. Press (2004).

Index