Methods of algebraic 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. Algebraic manifold, Projective manifold

Abstract. This is an introduction to algebraic geometry, with all the needed preliminaries included, and with emphasis on computations and applications. We first discuss plane geometry and curves, starting from the ancient Greeks, notably with results about conics, cubics, then spirals and lemniscates, and with the theory of the disciminant and the Cardano formulae explained too. Then we have a similar discussion in space, dealing this time with both curves and surfaces, and involving some abstract algebra too. We then discuss the study of the general algebraic manifolds, using abstract algebra, and their main properties. We end with an introduction to projective geometry.

Preface

How old is algebraic geometry? As old as mathematics itself, as known by us humans, because whenever you realize that a centered circle in the plane is given by $x^2 + y^2 = R^2$, that is already good algebraic geometry. In fact, there is no even need for coordinates for having some algebraic geometry started, because the old theorems of Thales, Pythagoras, Pappus and the other ancients are certainly quality algebraic geometry too.

At a more advanced level, you have ellipses and other conics, whose study can be quite tricky, followed by cubics and higher degree curves, such as spirals, lemniscates, and many other interesting beasts, whose study, again, requires an excellent knowledge of trigonometry, and plane geometry in general. In addition, there is a lot of interesting underlying physics in all this, with the ellipses and other conics being known from Kepler and Newton to be the trajectories of objects under the influence of gravity, and with more complicated curves appearing for instance as field lines, in gravity, or electrostatics.

And things are not over here with the plane curves, because if you try to intersect two such curves, say with one of them chosen to be a line, for simplifying things, that will lead you into the question of finding the roots of polynomials $P \in \mathbb{R}[X]$. And with this being a well-known subtle question, depending on the degree $n = \deg P$, with here $n = 2$ being known to be easy, but still watch out for $n = 2$ tricks that I might know, and that you don't, then with $n = 3, 4$ being the business of the theory of the resultant and discriminant, and the Cardano formulae, which are quite tricky mathematics, and then with $n = 5$ and higher requiring some help from people like Galois.

Passed the plane, we have the space, and here things ramify, because we can look either at curves, or at surfaces, with both being natural generalizations of the plane curves. And again, lots of things are known here, since ages, of varying levels of difficulty, and with everything needing, in addition, a bit of abstract algebra, for proper understanding.

In short, all sorts of interesting mathematics to be learned, and please don't listen to that graduate algebraic geometry student friend of yours, who might have told you that algebraic geometry was born in the 20th century. That is plainly wrong, the 20th century theory only comes after mastering the curves and surfaces, this is how things go. Actually, in order to see that your nerd friend does not know what he's talking about,

present him a good geometry problem, Greek style, with lots of lines and circles, and ask him for a quick solution, using his knowledge of schemes, and other modern beasts. You will be surprised, and can even have some fun of him, if priorly reading this book.

Which brings us to the present book. This book, and you guessed it right, is an introduction to algebraic geometry, with all the needed preliminaries included, and with emphasis on old style mathematics, concrete things, computations and applications. Modernity will be not forgotten, either, with an introduction to it, at the end of the book.

More in detail, the book is organized in four parts, as follows:

(1) We first discuss plane geometry and curves, starting from the ancient Greeks, notably with results about conics, cubics, then spirals, lemniscates and more.

(2) Then we have a similar discussion in space, dealing this time with both curves and surfaces, and involving some abstract algebra too.

(3) We then discuss the study of general algebraic manifolds, using abstract algebra, and their main properties, following Noether, Hilbert, Zariski and others.

(4) We end with an introduction to projective algebraic geometry, and to more advanced aspects of algebraic geometry, ancient or modern, in general.

Many thanks to my parents, both algebraic geometers, and with my father being particularly passioned by his work. Life was quite special for us kids, with the first four words that I learned in life being Mom, Dad, Fiber bundle and Grothendieck. We went on both kids to do mathematics too, but rather quantum mechanics, messing up things there, me in relation with the Heisenberg approach, and with my sister Valeria doing Schrödinger. Finally, many thanks to my cats, with my daily job of quantum physicist, every now and then I forget some algebra basics, but cats know all this well.

Cergy, March 2025
Teo Banica

Contents

Part I

Geometry, curves

Cheri, cheri lady
Going through a motion
Love is where you find it
Listen to your heart

CHAPTER 1

Plane 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 5.1, 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 $K \not|| L$.

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 $K \not|| L$ determine a point, $P = K \cap L$.*

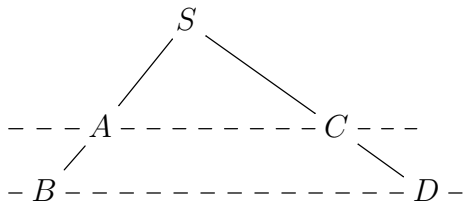
Very good, and now with Axiom 5.1 and Axiom 1.8 in hand, we are potentially ready for doing some geometry. However, this is not exactly true, and we will need as well:

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 $AC||BD$.

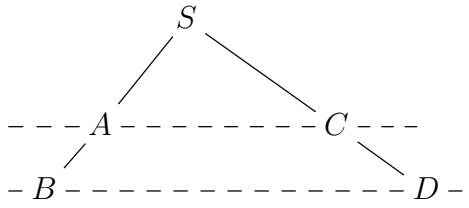
PROOF. We have indeed the following computation, based on the usual area formula for the triangles, that is, half of side times height, used multiple times:

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

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,*

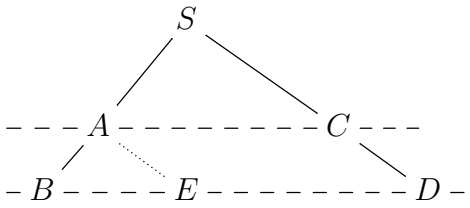


the following equality, involving the same number, holds as well:

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

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:



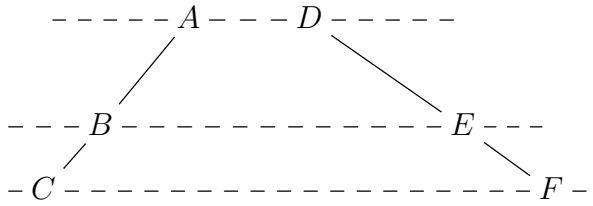
By using Theorem 1.10, we have then the following computation, as desired:

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

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:

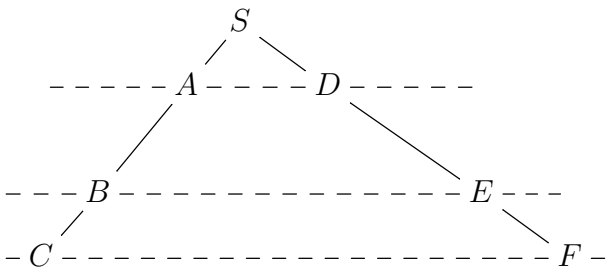
$$\frac{AB}{BC}=\frac{DE}{EF}$$

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.

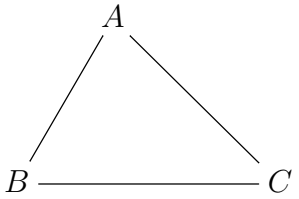
1b. Angles, triangles

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:

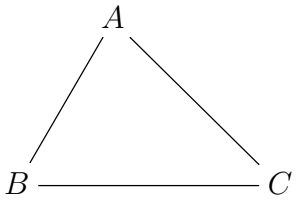
(1) Come with a small circle, inside ABC, and then inflate it, as to touch all 3 edges. The center of the circle will be then at equal distance from all 3 edges, so it will lie on all 3 angle bisectors. Thus, we have constructed the incenter, as required.

(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:

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

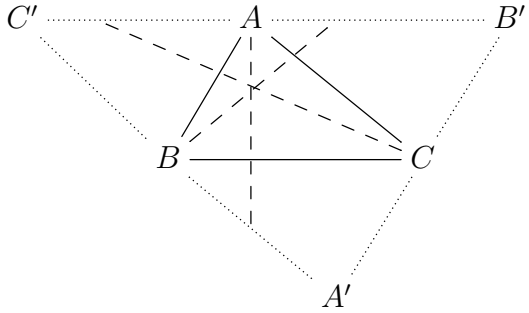
Thus P lies on the median emanating from A, and a similar argument shows that P lies as well on the medians emanating from B, C. Thus, we have our barycenter.

(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:

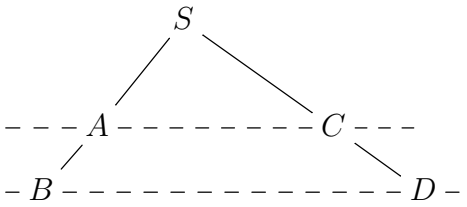


Thus, we are led to the conclusions in the statement. □

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:

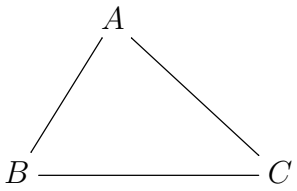
DEFINITION 1.16. *We can talk about the numeric value of angles, as follows:*

(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.

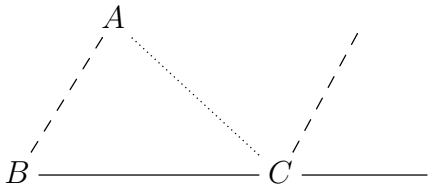
Getting back to work now, theorems and proofs, in relation with the above, here is a key result, which will be our main tool for the study of the angles:

THEOREM 1.17. *In an arbitrary triangle*



the sum of all three angles is 180°.

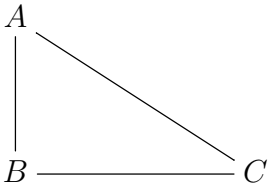
PROOF. This does not seem obvious to prove, with bare hands, but as usual, in such situations, some tricky parallels can come to the rescue. Let us prolong indeed the segment BC a bit, on the C side, and then draw a parallel at C, to the line AB, as follows:



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:

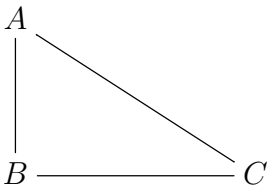
DEFINITION 1.18. *We call right triangle a triangle of type*



having one of the angles equal to 90°.

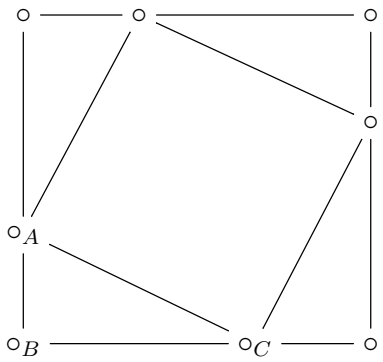
Many things can be said about right triangles, in particular with:

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:

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

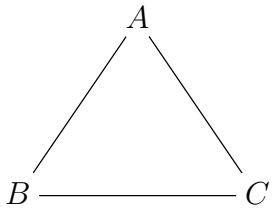
On the other hand, the outer square is made of the smaller square, having side AC, and of four identical right triangles, having sizes AB, BC. Thus:

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

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,*

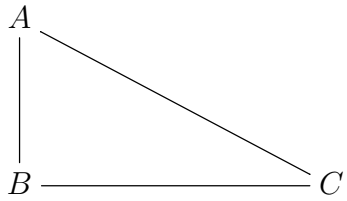


all angles equal 60°.

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

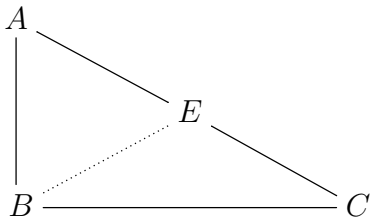
Another interesting angle is the 30° one. About it, we have:

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. □

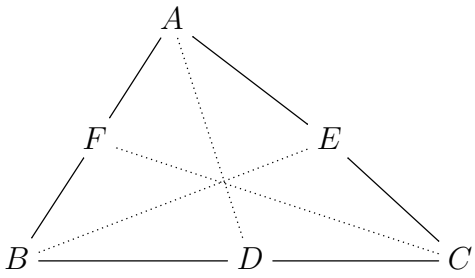
We have as well the following result, going back in time, to Pappus:

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. □

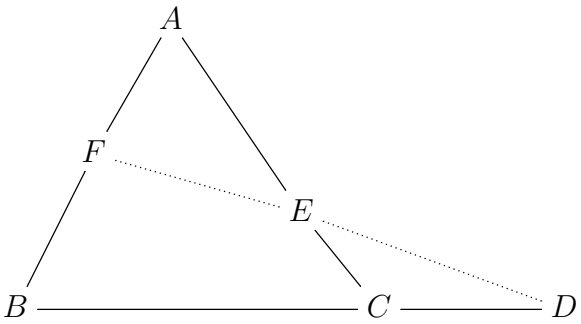
Let us go back now to basic triangle geometry and centers, as developed before in this chapter. In order to further build on that material, and systematically look at triangle

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:

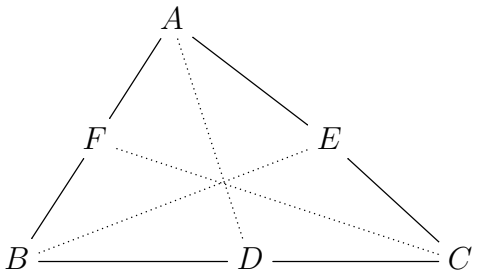
$$\frac{AF}{FB}\cdot\frac{BD}{DC}\cdot\frac{CE}{EA}=-1$$

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. □

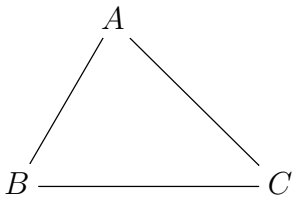
As a basic application of the Ceva theorem, we have now a new point of view on the barycenter. Indeed, the fact that the medians of a triangle cross can be seen as coming from the Ceva theorem, via the following trivial computation:

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

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.

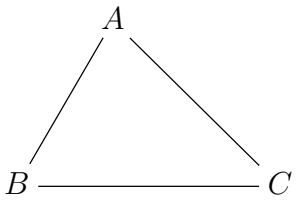
(2) Needless to say, in order for that Euler line to exist, as defined above, the triangle ABC must be assumed to be not equilateral. As for the basic example, for this, for an isosceles triangle, not equilateral, the Euler line is of course the symmetry axis.

(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 goal is that of having your own version of quantum field theory.

1d. Projective geometry

Switching topics, but still in relation with the parallel lines, that we constantly met in the above, you might have heard or not of projective geometry. In case you didn't yet, the general principle is that "this is the wonderland where parallel lines cross".

Which might sound a bit crazy, and not very realistic, but take a picture of some railroad tracks, and look at that picture. Do that parallel railroad tracks cross, on the picture? Sure they do. So, we are certainly not into abstractions here. QED.

Mathematically now, here are some axioms, to start with:

DEFINITION 1.28. *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 a basic example we have the usual projective plane $P_{\mathbb{R}}^2$, which is best seen as being the space of lines in $\mathbb{R}^3$ passing through the origin. To be more precise, let us call each of these lines in $\mathbb{R}^3$ passing through the origin a "point" of $P_{\mathbb{R}}^2$, and let us also call each plane in $\mathbb{R}^3$ passing through the origin a "line" of $P_{\mathbb{R}}^2$. Now observe the following:

(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.

Thus, what we have is a projective space in the sense of Definition 1.28. More generally now, we have the following construction, in arbitrary dimensions:

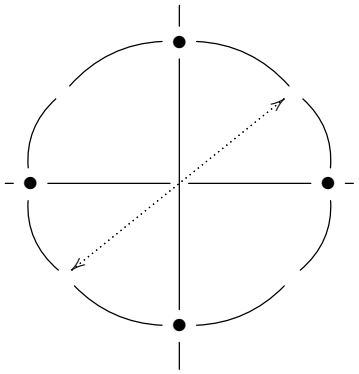
THEOREM 1.29. *We can define the projective space $P_{\mathbb{R}}^{N-1}$ as being the space of lines in $\mathbb{R}^N$ passing through the origin, and in small dimensions:*

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

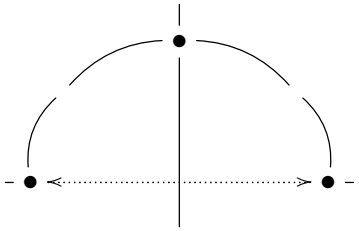
PROOF. We have several assertions here, with all this being of course a bit informal, and self-explanatory, the idea and some further details being as follows:

(1) To start with, the fact that the space $P_{\mathbb{R}}^{N-1}$ constructed in the statement is indeed a projective space in the sense of Definition 1.28 follows from definitions, exactly as in the discussion preceding the statement, regarding the case $N = 3$.

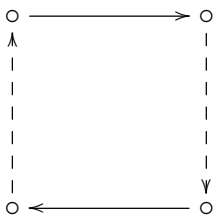
(2) At $N = 2$ now, a line in $\mathbb{R}^2$ passing through the origin corresponds to 2 opposite points on the unit circle $\mathbb{T} \subset \mathbb{R}^2$, according to the following scheme:



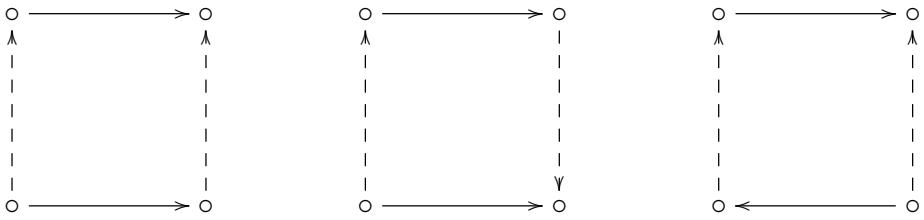
Thus, $P^1_{\mathbb{R}}$ corresponds to the upper semicircle of $\mathbb{T}$, with the endpoints identified, and so we obtain a circle, $P^1_{\mathbb{R}} = \mathbb{T}$, according to the following scheme:



(3) At $N = 3$, the space $P^2_{\mathbb{R}}$ 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:



(4) 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.

(5) 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.

(6) 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.

(7) 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 figure out 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.

All this is very interesting, but we will pause our study here, because we still have many other things to say. Getting now to finite fields, we have:

THEOREM 1.30. *Given a field F, we can talk about the projective space P^{N-1}_F, as being the space of lines in F^N passing through the origin. At $N = 3$ we have*

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

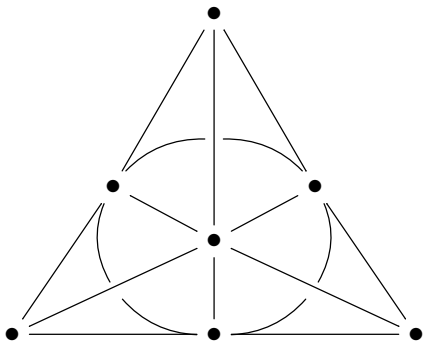
where $q = |F|$, in the case where our field F is finite.

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

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

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 $F = \mathbb{Z}_2$. 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 in Definition 1.28 are satisfied, in the sense that any two points determine a line, and any two lines determine a point. And isn't this beautiful.

1e. Exercises

Exercises:

EXERCISE 1.31.

EXERCISE 1.32.

EXERCISE 1.33.

EXERCISE 1.34.

EXERCISE 1.35.

EXERCISE 1.36.

EXERCISE 1.37.

EXERCISE 1.38.

Bonus exercise.

CHAPTER 2

Ellipses, conics

2a. Ellipses

Looking up, things are quite fascinating. The first thing that you see is the Sun, seemingly moving around the Earth on a circle, but a more careful study reveals that this circle is rather a deformed circle, called ellipsis. As for the other stars and planets, these have all sort of weird trajectories, but a more careful study reveals that, with due attention to what the best "center" is, the trajectories are often ellipses:

(1) Indeed, this applies to all the planets in our Solar System, which move around the biggest object in the system, which is by far the Sun, on ellipses.

(2) The same trick applies to the trajectories of various distant stars, the rule being always the same, "small moves around big, on an ellipsis".

(3) However, there are counterexamples too, such as asteroids reaching our Solar system, but then travelling outwards, never to be seen again.

Summarizing, modulo some annoying asteroids that we will leave for later, we are led in this way to ellipses, and their mathematics. And good news, a full theory of ellipses is available, and this since the ancient Greeks, whose main findings were as follows:

THEOREM 2.1. *The ellipses, taken centered at the origin* 0*, and squarely oriented with respect to* Oxy*, can be defined in* 4 *possible ways, as follows:*

(1) *As the curves given by an equation as follows, with* $a, b > 0$*:*

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

(2) *Or given by an equation as follows, with* $q > 0$*,* $p = -q$*, and* $l \in (0, 2q)$*:*

$$d(z, p) + d(z, q) = l$$

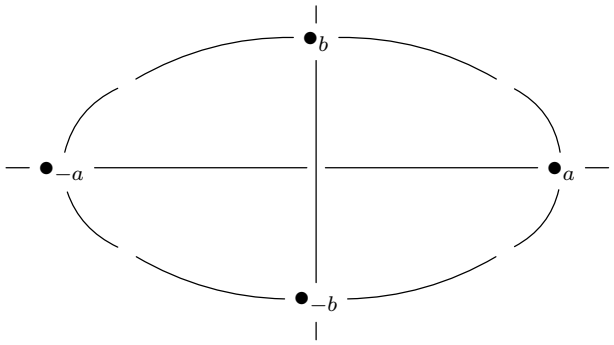
(3) *As the curves appearing when drawing a circle, from various perspectives:*

$$\bigcirc \quad \to \quad ?$$

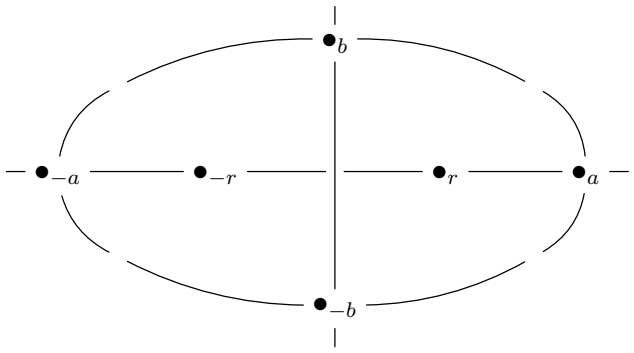
(4) *As the closed non-degenerate curves appearing by cutting a cone with a plane.*

PROOF. This might look a bit confusing, and you might say, what exactly is to be proved here. Good point, and in answer, what is to be proved is that the above constructions (1-4) give rise to the same class of curves. And this can be done as follows:

(1) To start with, let us draw a picture from what comes out of (1), which will be our main definition for the ellipses, in what follows. Here that is, making it clear what the parameters $a, b > 0$ stand for, with $2a \times 2b$ being the gift box size for our ellipsis:



(2) Let us prove now that such an ellipsis has two focal points, as stated in (2). We must look for a number $r > 0$, and a number $l > 0$, such that our ellipsis appears as $d(z, p) + d(z, q) = l$, with $p = (0, -r)$ and $q = (0, r)$, according to the following picture:



(3) Let us first compute these numbers $r, l > 0$. Assuming that our result holds indeed as stated, by taking $z = (0, a)$, we see that the length l is:

$$l = (a - r) + (a + r) = 2a$$

As for the parameter r, by taking $z = (b, 0)$, we conclude that we must have:

$$2\sqrt{b^2 + r^2} = 2a \implies r = \sqrt{a^2 - b^2}$$

(4) With these observations made, let us prove now the result. Given $l, r > 0$, and setting $p = (0, -r)$ and $q = (0, r)$, we have the following computation, with $z = (x, y)$:

$$\begin{aligned}
& d(z,p) + d(z,q) = l \\
\iff \quad & \sqrt{(x+r)^2 + y^2} + \sqrt{(x-r)^2 + y^2} = l \\
\iff \quad & \sqrt{(x+r)^2 + y^2} = l - \sqrt{(x-r)^2 + y^2} \\
\iff \quad & (x+r)^2 + y^2 = (x-r)^2 + y^2 + l^2 - 2l\sqrt{(x-r)^2 + y^2} \\
\iff \quad & 2l\sqrt{(x-r)^2 + y^2} = l^2 - 4xr \\
\iff \quad & 4l^2(x^2 + r^2 - 2xr + y^2) = l^4 + 16x^2r^2 - 8l^2xr \\
\iff \quad & 4l^2x^2 + 4l^2r^2 + 4l^2y^2 = l^4 + 16x^2r^2 \\
\iff \quad & (4x^2 - l^2)(4r^2 - l^2) = 4l^2y^2
\end{aligned}$$

(5) Now observe that we can further process the equation that we found as follows:

$$\begin{aligned}
(4x^2 - l^2)(4r^2 - l^2) = 4l^2y^2 \quad & \iff \quad \frac{4x^2 - l^2}{l^2} = \frac{4y^2}{4r^2 - l^2} \\
& \iff \quad \frac{4x^2 - l^2}{l^2} = \frac{y^2}{r^2 - l^2/4} \\
& \iff \quad \left(\frac{x}{2l}\right)^2 - 1 = \left(\frac{y}{\sqrt{r^2 - l^2/4}}\right)^2 \\
& \iff \quad \left(\frac{x}{2l}\right)^2 + \left(\frac{y}{\sqrt{r^2 - l^2/4}}\right)^2 = 1
\end{aligned}$$

(6) Thus, our result holds indeed, and with the numbers $l, r > 0$ appearing, and no surprise here, via the formulae $l = 2a$ and $r = \sqrt{a^2 - b^2}$, found in (3) above.

(7) Getting back now to our theorem, we have two other assertions there at the end, labeled (3,4). But, thinking a bit, these assertions are in fact equivalent, and in what concerns us, we will rather focus on (4), which looks more mathematical.

(8) And in what regards this assertion (4), this can be established indeed, by doing some 3D computations, that we will leave here as an instructive exercise. And with the promise that we will come back to this in a moment, with a full proof, in a more general setting, that of the conics, including the parabolas and hyperbolas as well. □

Many other things can be said about ellipses, as for instance about their writing in polar coordinates. We will be back to this, after introducing the other conics too.

2b. Further conics

All this is very nice, but before getting into physics, with some explanations for the fact that planets travel indeed on ellipses, which is something that we must surely understand, let us settle as well the question of wandering asteroids.

Observations show that these can travel on parabolas and hyperbolas, so what we need as mathematics is a unified theory of ellipses, parabolas and hyperbolas.

And fortunately, this theory exists, also since the ancient Greeks, as follows:

THEOREM 2.2. *The conics, which are the algebraic curves of degree* 2 *in the plane,*

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

with $\deg P \leq 2$*, appear modulo degeneration by cutting a* 2*-sided cone with a plane, and can be classified into ellipses, parabolas and hyperbolas.*

PROOF. This follows by further building on Theorem 2.1, as follows:

(1) Let us first classify the conics up to non-degenerate linear transformations of the plane, which are by definition transformations as follows, with $\det A \neq 0$:

$$\begin{pmatrix} x \\ y \end{pmatrix} \to A \begin{pmatrix} x \\ y \end{pmatrix}$$

Our claim is that as solutions we have the circles, parabolas, hyperbolas, along with some degenerate solutions, namely $\emptyset$, points, lines, pairs of lines, $\mathbb{R}^2$.

(2) As a first remark, it looks like we forgot precisely the ellipses, but via linear transformations these become circles, so things fine. As a second remark, all our claimed solutions can appear. Indeed, the circles, parabolas, hyperbolas can appear as follows:

$$x^2 + y^2 = 1 \quad , \quad x^2 = y \quad , \quad xy = 1$$

As for $\emptyset$, points, lines, pairs of lines, $\mathbb{R}^2$, these can appear too, as follows, and with our polynomial P chosen, whenever possible, to be of degree exactly 2:

$$x^2 = -1 \quad , \quad x^2 + y^2 = 0 \quad , \quad x^2 = 0 \quad , \quad xy = 0 \quad , \quad 0 = 0$$

Observe here that, when dealing with these degenerate cases, assuming $\deg P = 2$ instead of $\deg P \leq 2$ would only rule out $\mathbb{R}^2$ itself, which is not worth it.

(3) Getting now to the proof of our claim in (1), classification up to linear transformations, consider an arbitrary conic, written as follows, with $a, b, c, d, e, f \in \mathbb{R}$:

$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

Assume first $a \neq 0$. By making a square out of ax^2, up to a linear transformation in (x, y), we can get rid of the term cxy, and we are left with:

$$ax^2 + by^2 + dx + ey + f = 0$$

In the case $b \neq 0$ we can make two obvious squares, and again up to a linear transformation in (x, y), we are left with an equation as follows:

$$x^2 \pm y^2 = k$$

In the case of positive sign, $x^2 + y^2 = k$, the solutions are the circle, when $k \geq 0$, the point, when $k = 0$, and $\emptyset$, when $k < 0$. As for the case of negative sign, $x^2 - y^2 = k$, which reads $(x-y)(x+y) = k$, here once again by linearity our equation becomes $xy = l$, which is a hyperbola when $l \neq 0$, and two lines when $l = 0$.

(4) In the case $b \neq 0$ the study is similar, with the same solutions, so we are left with the case $a = b = 0$. Here our conic is as follows, with $c, d, e, f \in \mathbb{R}$:

$$cxy + dx + ey + f = 0$$

If $c \neq 0$, by linearity our equation becomes $xy = l$, which produces a hyperbola or two lines, as explained before. As for the remaining case, $c = 0$, here our equation is:

$$dx + ey + f = 0$$

But this is generically the equation of a line, unless we are in the case $d = e = 0$, where our equation is $f = 0$, having as solutions $\emptyset$ when $f \neq 0$, and $\mathbb{R}^2$ when $f = 0$.

(5) Thus, done with the classification, up to linear transformations as in (1). But this classification leads to the classification in general too, by applying now linear transformations to the solutions that we found. So, done with this, and very good.

(6) It remains to discuss the cone cutting. By suitably choosing our coordinate axes (x, y, z), we can assume that our cone is given by an equation as follows, with $k > 0$:

$$x^2 + y^2 = kz^2$$

In order to prove the result, we must in principle intersect this cone with an arbitrary plane, which has an equation as follows, with $(a, b, c) \neq (0, 0, 0)$:

$$ax + by + cz = d$$

(7) However, before getting into computations, observe that what we want to find is a certain degree 2 equation in the above plane, for the intersection. Thus, it is convenient to change the coordinates, as for our plane to be given by the following equation:

$$z = 0$$

(8) But with this done, what we have to do is to see how the cone equation $x^2+y^2 = kz^2$ changes, under this change of coordinates, and then set $z = 0$, as to get the (x, y) equation of the intersection. But this leads, via some thinking or computations, to the conclusion

that the cone equation $x^2+y^2=kz^2$ becomes in this way a degree 2 equation in (x,y), which can be arbitrary, and so to the final conclusion in the statement. □

2c. Gravity basics

Ready for some physics? All we have so far is certainly nice, but a bit too mathematical, seemingly away from the real life, and its problems. To be more precise, browsing through what we did so far, we can basically count on two applications of that:

(1) First we have Theorem 2.1 (3), teaching us how to draw circles, from different perspectives. But that rather belongs to Art and Humanities.

(2) We have as well Theorem 2.1 (4), teaching us that when coming with a big Viking axe, and slashing a conical tree stump, we get an ellipsis. Interesting too.

In short, and hope you get my point, before going ahead with more math, let us make sure that what we're doing is relevant to physics.

As a first question, we could try to understand how objects fall, under the influence of uniform gravity. And here we obtain parabolic trajectories.

2d. Kepler and Newton

At a more advanced level, of usual, non-uniform gravity, good news, not only what we did with conics is relevant, but is actually at the origin of modern physics, thanks to:

THEOREM 2.3. *Planets and other celestial bodies move around the Sun on conics,*

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

with $P\in\mathbb{R}[x,y]$ *being of degree* 2*, which can be ellipses, parabolas or hyperbolas.*

PROOF. This is something quite long, due to Kepler and Newton, but no fear, we know calculus, and threrefore what can resist us. The proof goes as follows:

(1) According to observations and calculations performed over the centuries, since the ancient times, and first formalized by Newton, following some groundbreaking work of Kepler, the force of attraction between two bodies of masses M,m is given by:

$$||F||=G\cdot\frac{Mm}{d^2}$$

Here d is the distance between the two bodies, and $G\simeq 6.674\times 10^{-11}$ is a constant. Now assuming that M is fixed at $0\in\mathbb{R}^3$, the force exterted on m positioned at $x\in\mathbb{R}^3$,

regarded as a vector $F \in \mathbb{R}^3$, is given by the following formula:

$$\begin{aligned} F &= -||F|| \cdot \frac{x}{||x||} \\ &= -\frac{GMm}{||x||^2} \cdot \frac{x}{||x||} \\ &= -\frac{GMmx}{||x||^3} \end{aligned}$$

But $F = ma = m\ddot{x}$, with $a = \ddot{x}$ being the acceleration, second derivative of the position, so the equation of motion of m, assuming that M is fixed at 0, is:

$$\ddot{x} = -\frac{GMx}{||x||^3}$$

Obviously, the problem happens in 2 dimensions, and you can even find, as an exercise, a formal proof of that, based on the above equation, if your really want to. Now here the most convenient is to use standard x, y coordinates, and denote our point as $z = (x, y)$. With this change made, and by setting $K = GM$, the equation of motion becomes:

$$\ddot{z} = -\frac{Kz}{||z||^3}$$

(2) The idea now is that the problem can be solved via some calculus. Let us write indeed our vector $z = (x, y)$ in polar coordinates, as follows:

$$x = r\cos\theta \quad , \quad y = r\sin\theta$$

We have then $||z|| = r$, and our equation of motion becomes:

$$\ddot{z} = -\frac{Kz}{r^3}$$

Let us differentiate now x, y. By using the standard calculus rules, we have:

$$\dot{x} = \dot{r}\cos\theta - r\sin\theta \cdot \dot{\theta}$$

$$\dot{y} = \dot{r}\sin\theta + r\cos\theta \cdot \dot{\theta}$$

Differentiating one more time gives the following formulae:

$$\ddot{x} = \ddot{r}\cos\theta - 2\dot{r}\sin\theta \cdot \dot{\theta} - r\cos\theta \cdot \dot{\theta}^2 - r\sin\theta \cdot \ddot{\theta}$$

$$\ddot{y} = \ddot{r}\sin\theta + 2\dot{r}\cos\theta \cdot \dot{\theta} - r\sin\theta \cdot \dot{\theta}^2 + r\cos\theta \cdot \ddot{\theta}$$

Consider now the following two quantities, appearing as coefficients in the above:

$$a = \ddot{r} - r\dot{\theta}^2 \quad , \quad b = 2\dot{r}\dot{\theta} + r\ddot{\theta}$$

In terms of these quantities, our second derivative formulae read:

$$\ddot{x} = a\cos\theta - b\sin\theta$$

$$\ddot{y} = a\sin\theta + b\cos\theta$$

(3) We can now solve the equation of motion from (1). Indeed, with the formulae that we found for $\ddot{x}, \ddot{y}$, our equation of motion takes the following form:

$$a\cos\theta - b\sin\theta = -\frac{K}{r^2}\cos\theta$$

$$a\sin\theta + b\cos\theta = -\frac{K}{r^2}\sin\theta$$

But these two formulae can be written in the following way:

$$\left(a + \frac{K}{r^2}\right)\cos\theta = b\sin\theta \quad , \quad \left(a + \frac{K}{r^2}\right)\sin\theta = -b\cos\theta$$

By making now the product, and assuming that we are in a non-degenerate case, where the angle θ varies indeed, we obtain by positivity that we must have:

$$a + \frac{K}{r^2} = b = 0$$

(4) Let us first examine the second equation, $b = 0$. This can be solved as follows:

$$\begin{aligned} b = 0 &\iff 2\dot{r}\dot{\theta} + r\ddot{\theta} = 0\\ &\iff \frac{\ddot{\theta}}{\dot{\theta}} = -2\frac{\dot{r}}{r}\\ &\iff (\log\dot{\theta})' = (-2\log r)'\\ &\iff \log\dot{\theta} = -2\log r + c\\ &\iff \dot{\theta} = \frac{\lambda}{r^2} \end{aligned}$$

As for the first equation the we found, namely $a + K/r^2 = 0$, this becomes:

$$\ddot{r} - \frac{\lambda^2}{r^3} + \frac{K}{r^2} = 0$$

As a conclusion to all this, in polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$, our equations of motion are as follows, with λ being a constant, not depending on t:

$$\ddot{r} = \frac{\lambda^2}{r^3} - \frac{K}{r^2} \quad , \quad \dot{\theta} = \frac{\lambda}{r^2}$$

Even better now, by writing $K = \lambda^2/c$, these equations read:

$$\ddot{r} = \frac{\lambda^2}{r^2}\left(\frac{1}{r} - \frac{1}{c}\right) \quad , \quad \dot{\theta} = \frac{\lambda}{r^2}$$

(5) In order to study the first equation, we use a trick. Let us write:

$$r(t) = \frac{1}{f(\theta(t))}$$

Abbreviated, and by reminding that f takes $\theta = \theta(t)$ as variable, this reads:

$$r = \frac{1}{f}$$

With the convention that dots mean as usual derivatives with respect to t, and that the primes will denote derivatives with respect to $\theta = \theta(t)$, we have:

$$\dot{r} = -\frac{f'\dot{\theta}}{f^2} = -\frac{f'}{f^2} \cdot \frac{\lambda}{r^2} = -\lambda f'$$

By differentiating one more time with respect to t, we obtain:

$$\ddot{r} = -\lambda f''\dot{\theta} = -\lambda f'' \cdot \frac{\lambda}{r^2} = -\frac{\lambda^2}{r^2} f''$$

On the other hand, our equation for $\ddot{r}$ found in (4) above reads:

$$\ddot{r} = \frac{\lambda^2}{r^2}\left(\frac{1}{r} - \frac{1}{c}\right) = \frac{\lambda^2}{r^2}\left(f - \frac{1}{c}\right)$$

Thus, in terms of $f = 1/r$ as above, our equation for $\ddot{r}$ simply reads:

$$f'' + f = \frac{1}{c}$$

But this latter equation is elementary to solve. Indeed, both functions $\cos t, \sin t$ satisfy $g" + g = 0$, so any linear combination of them satisfies as well this equation. But the solutions of $f'' + f = 1/c$ being those of $g'' + g = 0$ shifted by $1/c$, we obtain:

$$f = \frac{1 + \varepsilon\cos\theta + \delta\sin\theta}{c}$$

Now by inverting, we obtain the following formula:

$$r = \frac{c}{1 + \varepsilon\cos\theta + \delta\sin\theta}$$

(6) But this leads to the conclusion that the trajectory is a conic. Indeed, in terms of the parameter θ, the formulae of the coordinates are:

$$x = \frac{c\cos\theta}{1 + \varepsilon\cos\theta + \delta\sin\theta} \quad , \quad y = \frac{c\sin\theta}{1 + \varepsilon\cos\theta + \delta\sin\theta}$$

Now observe that these two functions x, y satisfy the following formula:

$$x^2 + y^2 = \frac{c^2(\cos^2\theta + \sin^2\theta)}{(1 + \varepsilon\cos\theta + \delta\sin\theta)^2} = \frac{c^2}{(1 + \varepsilon\cos\theta + \delta\sin\theta)^2}$$

On the other hand, these two functions satisfy as well the following formula:

$$\begin{aligned}(\varepsilon x+\delta y-c)^2 &= \frac{c^2\big(\varepsilon\cos\theta+\delta\sin\theta-(1+\varepsilon\cos\theta+\delta\sin\theta)\big)^2}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}\\ &= \frac{c^2}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2}\end{aligned}$$

We conclude that our coordinates x,y satisfy the following equation:

$$x^2+y^2=(\varepsilon x+\delta y-c)^2$$

But what we have here is an equation of a conic, and we are done. □

We have the following result, coming as a useful version of Theorem 2.3:

THEOREM 2.4. *In the context of a 2-body problem, with M fixed at 0, and m starting its movement from Ox, the equation of motion of m, namely*

$$\ddot{z}=-\frac{Kz}{||z||^3}$$

with $K=GM$, and $z=(x,y)$, becomes in polar coordinates, $x=r\cos\theta$, $y=r\sin\theta$,

$$\ddot{r}=\frac{\lambda^2}{r^2}\left(\frac{1}{r}-\frac{1}{c}\right)\quad,\quad\dot{\theta}=\frac{\lambda}{r^2}$$

for some $\lambda,c\in\mathbb{R}$, related by $\lambda^2=Kc$. The value of r in terms of θ is given by

$$r=\frac{c}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

for some $\varepsilon,\delta\in\mathbb{R}$. At the level of the affine coordinates x,y, this means

$$x=\frac{c\cos\theta}{1+\varepsilon\cos\theta+\delta\sin\theta}\quad,\quad y=\frac{c\sin\theta}{1+\varepsilon\cos\theta+\delta\sin\theta}$$

with $\theta=\theta(t)$ being subject to $\dot{\theta}=\lambda^2/r$, as above. Finally, we have

$$x^2+y^2=(\varepsilon x+\delta y-c)^2$$

which is a degree 2 equation, and so the resulting trajectory is a conic.

PROOF. This is a sort of "best of" the formulae found in the proof of Theorem 2.3. And in the hope of course that we have not forgotten anything. □

There are of course many other things that can be said, as a continuation of the above. For instance, we would like to understand how the various motion parameters $\lambda,c,\varepsilon,\delta$ appear from the initial data. And the formulae here are as follows:

PROPOSITION 2.5. *In the context of Theorem 2.4, and in polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$, the initial data is as follows, with $R = r_0$:*

$$r_0 = \frac{c}{1+\varepsilon} \quad , \quad \theta_0 = 0$$

$$\dot{r}_0 = -\frac{\delta\sqrt{K}}{\sqrt{c}} \quad , \quad \dot{\theta}_0 = \frac{\sqrt{Kc}}{R^2}$$

$$\ddot{r}_0 = \frac{\varepsilon K}{R^2} \quad , \quad \ddot{\theta}_0 = \frac{4\delta K}{R^2}$$

The corresponding formulae for the affine coordinates x, y can be deduced from this. Also, the various motion parameters c, ε, δ and $\lambda = \sqrt{Kc}$ can be recovered from this data.

PROOF. We have several assertions here, the idea being as follows:

(1) As mentioned in Theorem 2.4, the object m begins its movement on Ox. Thus we have $\theta_0 = 0$, and from this we get the formula of r_0 in the statement.

(2) Regarding now the initial speed, the formula of $\dot{\theta}_0$ follows from:

$$\dot{\theta} = \frac{\lambda}{r^2} = \frac{\sqrt{Kc}}{r^2}$$

Also, in what concerns the radial speed, the formula of $\dot{r}_0$ follows from:

$$\dot{r} = \frac{c(\varepsilon\sin\theta - \delta\cos\theta)\dot{\theta}}{(1+\varepsilon\cos\theta+\delta\sin\theta)^2} = \frac{\sqrt{K}(\varepsilon\sin\theta-\delta\cos\theta)}{\sqrt{c}}$$

(3) Regarding now the initial acceleration, by using $\dot{\theta} = \sqrt{Kc}/r^2$ we find:

$$\ddot{\theta} = -2\sqrt{Kc}\cdot\frac{2r\dot{r}}{r^3} = -\frac{4\sqrt{Kc}\cdot\dot{r}}{r^2}$$

In particular at $t = 0$ we obtain the formula in the statement, namely:

$$\ddot{\theta}_0 = -\frac{4\sqrt{Kc}\cdot\dot{r}_0}{R^2} = \frac{4\sqrt{Kc}}{R^2}\cdot\frac{\delta\sqrt{K}}{\sqrt{c}} = \frac{4\delta K}{R^2}$$

(4) Also regarding acceleration, with $\lambda = \sqrt{Kc}$ our main motion formula reads:

$$\ddot{r} = \frac{Kc}{r^2}\left(\frac{1}{r}-\frac{1}{c}\right)$$

In particular at $t = 0$ we obtain the formula in the statement, namely:

$$\ddot{r}_0 = \frac{Kc}{R^2}\left(\frac{1}{R}-\frac{1}{c}\right) = \frac{Kc}{R^2}\cdot\frac{\varepsilon}{c} = \frac{\varepsilon K}{R^2}$$

(5) Finally, the last assertion is clear, and since the formulae look better anyway in polar coordinates than in affine coordinates, we will not get into details here. □

With the above formulae in hand, we can work out how various inital speeds and accelerations lead to various types of conics. The computations here are many, and very interesting, and we will leave them as a long, pleasant and instructive exercise.

2e. Exercises

Exercises:

Exercise 2.6.

Exercise 2.7.

Exercise 2.8.

Exercise 2.9.

Exercise 2.10.

Exercise 2.11.

Exercise 2.12.

Exercise 2.13.

Bonus exercise.

CHAPTER 3

Plane curves

3a. Plane curves

As a conclusion to what we did so far, conics are at the core of everything, mathematics, physics, life. But, what is next? A natural answer to this question comes from:

DEFINITION 3.1. *An algebraic curve in $\mathbb{R}^2$ is the vanishing set*

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

of a polynomial $P \in \mathbb{R}[X,Y]$ of arbitrary degree.

We already know well the algebraic curves in degree 2, which are the conics, and a first problem is, what results from what we learned about conics have a chance to be relevant to the arbitrary algebraic curves. And normally none, because the ellipses, parabolas and hyperbolas are obviously very particular curves, having very particular properties.

Let us record however a useful statement here, as follows:

PROPOSITION 3.2. *The conics can be written in cartesian, polar, parametric or complex coordinates, with the equations for the unit circle being*

$$x^2 + y^2 = 1 \quad , \quad r = 1 \quad , \quad x = \cos t \, , \, y = \sin t \quad , \quad |z| = 1$$

and with the equations for ellipses, parabolas and hyperbolas being similar.

PROOF. The equations for the circle are clear, those for ellipses can be found in the above, and we will leave as an exercise those for parabolas and hyperbolas. □

As a true answer to our question now, coming this time from a very modest conic, namely $xy = 0$, that we dismissed in the above as being "degenerate", we have:

THEOREM 3.3. *The following happen, for curves C defined by polynomials P:*

(1) *In degree $d = 2$, curves can have singularities, such as $xy = 0$ at $(0,0)$.*
(2) *In general, assuming $P = P_1 \ldots P_k$, we have $C = C_1 \cup \ldots\ldots \cup C_k$.*
(3) *A union of curves $C_i \cup C_j$ is generically non-smooth, unless disjoint.*
(4) *Due to this, we say that C is non-degenerate when P is irreducible.*

PROOF. All this is self-explanatory, the details being as follows:

(1) This is something obvious, just the story of two lines crossing.

(2) This comes from the following trivial fact, with the notation $z = (x, y)$:

$$P_1 \ldots P_k(z) = 0 \iff P_1(z) = 0, \text{ or } P_2(z) = 0, \ \ldots \ , \text{ or } P_k(z) = 0$$

(3) This is something very intuitive, and it actually takes a bit of time to imagine a situation where $C_1 \cap C_2 \neq \emptyset$, $C_1 \not\subset C_2$, $C_2 \not\subset C_1$, but $C_1 \cup C_2$ is smooth. In practice now, "generically" has of course a mathematical meaning, in relation with probability, and our assertion does say something mathematical, that we are supposed to prove. But, we will not insist on this, and leave this as an instructive exercise, precise formulation of the claim, and its proof, in the case you are familiar with probability theory.

(4) This is just a definition, based on the above, that we will use in what follows. □

With degree 1 and 2 investigated, and our conclusions recorded, let us get now to degree 3, see what new phenomena appear here. And here, to start with, we have the following remarkable curve, well-known from calculus, because 0 is not a maximum or minimum of the function $x \to y$, despite the derivative vanishing there:

$$x^3 = y$$

Also, in relation with set theory and logic, and with the foundations of mathematics in general, we have the following curve, which looks like the empyset $\emptyset$:

$$(x - y)(x^2 + y^2 - 1) = 0$$

But, it is not about counterexamples to calculus, or about logic, that we want to talk about here. As a first truly remarkable degree 3 curve, or cubic, we have the cusp:

PROPOSITION 3.4. *The standard cusp, which is the cubic given by*

$$x^3 = y^2$$

has a singularity at $(0, 0)$*, with only* 1 *tangent line at that singularity.*

PROOF. The two branches of the cusp are indeed both tangent to Ox, because:

$$y' = \pm\frac{3}{2}\sqrt{x} \implies y'(0) = 0$$

Observe also that what happens for the cusp is different from what happens for $xy = 0$, precisely because we have 1 line tangent at the singularity, instead of 2. □

As a second remarkable cubic, which gets the crown, and the right to have a Theorem about it, we have the Tschirnhausen curve, which is as follows:

THEOREM 3.5. *The Tschirnhausen cubic, given by the following equation,*

$$x^3 = x^2 - 3y^2$$

makes the dream of $xy = 0$ *come true, by self-intersecting, and being non-degenerate.*

PROOF. This is something self-explanatory, by drawing a picture, but there are several other interesting things that can be said about this curve, and the family of curves containing it, depending on a parameter, and up to basic transformations, as follows:

(1) Let us start with the curve written in polar coordinates as follows:

$$r\cos^3\left(\frac{\theta}{3}\right)=a$$

With $t=\tan(\theta/3)$, the equations of the coordinates are as follows:

$$x=a(1-3t^2)\quad,\quad y=at(3-t^2)$$

Now by eliminating t, we reach to the following equation:

$$(a-x)(8a+x)^2=27ay^2$$

(2) By translating horizontally by $8a$, and changing signs of variables, we have:

$$x=3a(3-t^2)\quad,\quad y=at(3-t^2)$$

Now by eliminating t, we reach to the following equation:

$$x^3=9a(x^2-3y^2)$$

But with $a=1/9$ this is precisely the equation in the statement. □

3b. Higher degree

In degree 4 now, quartics, we have enough dimensions for "improving" the cusp and the Tschirnhausen curve. First we have the cardioid, which is as follows:

PROPOSITION 3.6. *The cardioid, which is a quartic, given in polar coordinates by*

$$2r=a(1-\cos\theta)$$

makes the dream of $x^3=y^2$ come true, by being a closed curve, with a cusp.

PROOF. As before with the Tschirnhausen curve, this is something self-explanatory, by drawing a picture, but there are several things that must be said, as follows:

(1) The cardioid appears by definition by rolling a circle of radius $c>0$ around another circle of same radius $c>0$. With θ being the rolling angle, we have:

$$x=2c(1-\cos\theta)\cos\theta$$

$$y=2c(1-\cos\theta)\sin\theta$$

(2) Thus, in polar coordinates we get the equation in the statement, with $a=4c$:

$$r=2c(1-\cos\theta)$$

(3) Finally, in cartesian coordinates, the equation is as follows:

$$(x^2+y^2)^2+4cx(x^2+y^2)=4c^2y^2$$

Thus, what we have is indeed a degree 4 curve, as claimed. □

Still in degree 4, the crown gets to the Bernoulli lemniscate, which is as follows:

THEOREM 3.7. *The Bernoulli lemniscate, a quartic, which is given by*

$$r^2 = a^2 \cos 2\theta$$

makes the dream of $x^3 = x^2 - 3y^2$ *come true, by being closed, and self-intersecting.*

PROOF. As usual, this is something self-explanatory, by drawing a picture, which looks like ∞, but there are several other things that must be said, as follows:

(1) In cartesian coordinates, the equation is as follows, with $a^2 = 2c^2$:

$$(x^2 + y^2)^2 = c^2(x^2 - y^2)$$

(2) Also, we have the following nice complex reformulation of this equation:

$$|z + c| \cdot |z - c| = c^2$$

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

In degree 5, in the lack of any spectacular quintic, let us record:

THEOREM 3.8. *Unlike in degree* $3, 4$*, where equations can be solved, by the Cardano formula, in degree* 5 *this generically does not happen, an example being*

$$x^5 - x - 1 = 0$$

having Galois group S_5*, not solvable. Geometrically, this tells us that the intersection of the quintic* $y = x^5 - x - 1$ *with the line* $y = 0$ *cannot be computed.*

PROOF. Obviously off-topic, but with no good quintic available, and still a few more minutes before the bell ringing, I had to improvise a bit, and tell you about this:

(1) As indicated, the degree 3 equations can be solved a bit like the degree 2 ones, but with the formula, due to Cardano, being more complicated. With some square making tricks, which are non-trivial either, the Cardano formula applies to degree 4 as well.

(2) In degree 5 or higher, none of this is possible. Long story here, the idea being that in order for $P = 0$ to be solvable, the group $Gal(P)$ must be solvable, in the sense of group theory. But, unlike S_3, S_4 which are solvable, S_5 and higher are not solvable. □

Back now to our usual business, in degree 6, sextics, we first have here:

PROPOSITION 3.9. *The trefoil sextic, or Kiepert curve, which is given by*

$$r^3 = a^3 \cos 3\theta$$

looks like a trefoil, closed curve, with a triple self-intersection.

PROOF. As before, drawing a picture is mandatory. With $z = re^{i\theta}$ we have:

$$\begin{aligned}
r^3 = a^3 \cos 3\theta &\iff r^3 \cos 3\theta = \left(\frac{r^2}{a}\right)^3 \\
&\iff z^3 + \bar{z}^3 = 2\left(\frac{z\bar{z}}{a}\right)^3 \\
&\iff (x+iy)^3 + (x-iy)^3 = 2\left(\frac{x^2+y^2}{a}\right)^3 \\
&\iff x^3 - 3xy^2 = \left(\frac{x^2+y^2}{a}\right)^3 \\
&\iff (x^2+y^2)^3 = a^3(x^3 - 3xy^2)
\end{aligned}$$

Thus, we have indeed a sextic, as claimed. □

We also have in degree 6 the most beautiful of curves them all, the Cayley sextic:

THEOREM 3.10. *The Cayley sextic, given in polar coordinates by*

$$r = a\cos^3\left(\frac{\theta}{3}\right)$$

makes the dream of everyone come true, by looking like a self-intersecting heart.

PROOF. As before, picture mandatory. With $z = re^{i\theta}$ and $u = z^{1/3}$ we have:

$$\begin{aligned}
r = a\cos^3\left(\frac{\theta}{3}\right) &\iff ar\cos^3\left(\frac{\theta}{3}\right) = r^2 \\
&\iff a\left(\frac{u+\bar{u}}{2}\right)^3 = r^2 \\
&\iff a(u^3 + \bar{u}^3 + 3u\bar{u}(u+\bar{u})) = 8r^2 \\
&\iff 3au\bar{u} \cdot \frac{u+\bar{u}}{2} = 4r^2 - ax \\
&\iff 27a^3r^6 \cdot \frac{r^2}{a} = (4r^2 - ax)^3 \\
&\iff 27a^2(x^2+y^2)^2 = (4x^2 + 4y^2 - ax)^3
\end{aligned}$$

Thus, we have indeed a sextic, as claimed. □

3c. Sinusoidal spirals

Quite remarkably, most of the above curves are sinusoidal spirals, in the following sense, and with actually the term "sinusoidal spiral" being a bit unfortunate:

THEOREM 3.11. *The sinusoidal spirals, which are as follows,*

$$r^n = a^n \cos n\theta$$

with $a \neq 0$ and $n \in \mathbb{Q} - \{0\}$, include the following curves:

(1) $n = -1$ *line.*
(2) $n = 1$ *circle,* $n = -1/2$ *parabola,* $n = -2$ *hyperbola.*
(3) $n = -3$ *Humbert cubic,* $n = -1/3$ *Tschirnhausen curve.*
(4) $n = 1/2$ *cardioid,* $n = 2$ *Bernoulli lemniscate.*
(5) $n = 3$ *Kiepert trefoil,* $n = 1/3$ *Cayley sextic.*

PROOF. We first have to prove that the sinusoidal spirals are indeed algebraic curves. But this is best done by using the complex coordinate $z = re^{i\theta}$, as follows:

$$\begin{aligned} r^n = a^n \cos n\theta &\iff r^n \cos n\theta = \left(\frac{r^2}{a}\right)^n \\ &\iff z^n + \bar{z}^n = 2\left(\frac{z\bar{z}}{a}\right)^n \\ &\iff (x+iy)^n + (x-iy)^n = 2\left(\frac{x^2+y^2}{a}\right)^n \end{aligned}$$

As a first observation now, in the case $n \in \mathbb{N}$ we can simply use the binomial formula, and we get an algebraic equation of degree $2n$, as follows:

$$\sum_{k=0}^{[n/2]} (-1)^k \binom{n}{2k} x^{n-2k} y^{2k} = \left(\frac{x^2+y^2}{a}\right)^n$$

In general, things are a bit more complicated, as shown for instance by our computation for the Cayley sextic. However, the same idea as there applies, and we are led in this way to the equation of an algebraic curve, as claimed. Regarding now the examples:

(1) At $n = -1$ the equation is as follows, producing a line:

$$r\cos\theta = a \iff x = a$$

(2) At $n = 1$ the equation is as follows, producing a circle:

$$r = a\cos\theta \iff r^2 = ax \iff x^2 + y^2 = ax$$

(3) At $n = -1/2$ the equation is as follows, producing a parabola:

$$a = r\cos^2(\theta/2) \iff r + x = 2a \iff y^2 = 4a(a-x)$$

(4) At $n = -2$ the equation is as follows, producing a hyperbola:

$$a^2 = r\cos^2 2\theta \iff a^2 = 2x^2 - r^2 \iff (x+y)(x-y) = a^2$$

(5) At $n=-3$ the equation is as follows, producing a curve with 3 components, which looks like some sort of "trivalent hyperbola", called Humbert cubic:

$$r^3\cos 3\theta=a^3\iff z^3+\bar{z}^3=2a^3\iff x^3-3xy^2=a^3$$

(6) As for the other curves, this follows from our various formulae above. □

Let us study now more in detail the sinusoidal spirals. We first have:

PROPOSITION 3.12. *The sinusoidal spirals, which with $z=x+iy$ are*

$$z^n+\bar{z}^n=2\left(\frac{z\bar{z}}{a}\right)^n$$

with $a\neq 0$ and $n\in\mathbb{Q}-\{0\}$, are as follows:

(1) *With $n=-m$, $m\in\mathbb{N}$, the equation is $z^m+\bar{z}^m=2a^m$, degree m.*
(2) *With $n=m$, $m\in\mathbb{N}$, the equation is $z^m+\bar{z}^m=2(z\bar{z}/a)^m$, degree $2m$.*
(3) *With $n=-1/m$, $m\in\mathbb{N}$, the equation is $(z^{1/m}+\bar{z}^{1/m})^m=2^m a$.*
(4) *With $n=1/m$, $m\in\mathbb{N}$, the equation is $(z^{1/m}+\bar{z}^{1/m})^m=2^m z\bar{z}/a$.*

PROOF. This is something self-explanatory, the details being as follows:

(1) With $n=-m$ and $m\in\mathbb{N}$ as in the statement, the equation is, as claimed:

$$z^{-m}+\bar{z}^{-m}=2\left(\frac{z\bar{z}}{a}\right)^{-m}\iff z^m+\bar{z}^m=2a^m$$

(2) This is an empty statement, just a matter of using the new variable $m=n$.

(3) With $n=-1/m$ and $m\in\mathbb{N}$ as in the statement, the equation is, as claimed:

$$\begin{aligned}z^{-1/m}+\bar{z}^{-1/m}=2\left(\frac{z\bar{z}}{a}\right)^{-1/m}&\iff z^{1/m}+\bar{z}^{1/m}=2a^{1/m}\\&\iff (z^{1/m}+\bar{z}^{1/m})^m=2^m a\end{aligned}$$

(4) With $n=1/m$ and $m\in\mathbb{N}$ as in the statement, the equation is, as claimed:

$$z^{1/m}+\bar{z}^{1/m}=2\left(\frac{z\bar{z}}{a}\right)^{1/m}\iff (z^{1/m}+\bar{z}^{1/m})^m=2^m\cdot\frac{z\bar{z}}{a}$$

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

Observe that in the fractionary cases, $n=\pm 1/m$, the equations in the above statement are not polynomial in x,y, unless at very small values of m. To be more precise:

(1) In the case $n=-1/m$, we certainly have at $m=1,2,3$ the $d=1$ line, $d=2$ parabola, and $d=3$ Tschirnhausen curve, but at $m=4$ things change, with the equation $(z^{1/4}+\bar{z}^{1/4})^4=16a$ being no longer polynomial in x,y, and requiring a further square operation to make it polynomial, and therefore leading to a curve of degree $d=8$.

(2) As for the case $n = 1/m$, this is more complicated, with the data that we have at $m = 1, 2, 3$, namely the $d = 2$ circle, $d = 3$ cardioid, and $d = 6$ Cayley sextic, being not very good, and with things getting even more complicated at $m = 4$ and higher.

In short, things quite complicated, and the general case, $n = \pm p/q$ with $p, q \in \mathbb{N}$, is certainly even more complicated. Instead of insisting on this, let us focus now on the simplest sinusoidal spirals that we have, namely those with $n = \pm m$, with $m \in \mathbb{N}$.

3d. Lemniscates

The point indeed is that the sinusoidal spirals with $n \in \mathbb{N}$ are also part of another remarkable family of plane algebraic curves, going back to Cassini, as follows:

THEOREM 3.13. *The polynomial lemniscates, which are as follows,*

$$|P(z)| = b^n$$

with $P \in \mathbb{C}[X]$ having n distinct roots, and $b > 0$, include the following curves:

(1) *The sinusoidal spirals with $n \in \mathbb{N}$, including the $n = 1$ circle, $n = 2$ Bernoulli lemniscate, and $n = 3$ Kiepert trefoil.*
(2) *The Cassini ovals, which are the quartics given by $|z + c| \cdot |z - c| = b^2$, covering too the Bernoulli lemniscate, appearing at $b = c$.*

PROOF. This is something quite self-explanatory, the details being as follows:

(1) Regarding the sinusoidal spirals with $n \in \mathbb{N}$, their equation is, with $a^n = 2c^n$:

$$\begin{aligned} z^n + \bar{z}^n = 2\left(\frac{z\bar{z}}{a}\right)^n &\iff c^n(z^n + \bar{z}^n) = (z\bar{z})^n \\ &\iff (z^n - c^n)(\bar{z}^n - c^n) = c^{2n} \\ &\iff |z^n - c^n| = c^n \end{aligned}$$

(2) Regarding the Cassini ovals, these correspond to the case where the polynomial $P \in \mathbb{C}[X]$ has degree 2, and we already know from the above that these cover the Bernoulli lemniscate. In general, the equation for the Cassini ovals is:

$$\begin{aligned} |z + c| \cdot |z - c| = b^2 &\iff |z^2 - c^2| = b^2 \\ &\iff (z^2 - c^2)(\bar{z}^2 - c^2) = b^4 \\ &\iff (z\bar{z})^2 - c^2(z^2 + \bar{z}^2) + c^4 = b^4 \\ &\iff (x^2 + y^2)^2 - c^2(x^2 - y^2) + c^4 = b^4 \\ &\iff (x^2 + y^2)^2 = c^2(x^2 - y^2) + b^4 - c^4 \end{aligned}$$

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

The polynomial lemniscates can be geometrically understood as follows:

THEOREM 3.14. *The equation $|P(z)| = b$ defining the polynomial lemniscates can be written as follows, in terms of the roots $c_1, \ldots, c_n$ of the polynomial P,*

$$\sqrt[n]{\prod_{k=1}^{n} |z - c_i|} = b$$

telling us that the geometric mean of the distances from z to the vertices of the polygon formed by $c_1, \ldots, c_n$ must be the constant $b > 0$.

PROOF. This is something self-explanatory, and as an illustration, let us work out the case of sinusoidal spirals with $n \in \mathbb{N}$. Here with $w = e^{2\pi i/n}$ we have:

$$z^n - c^n = \prod_{k=1}^{n} (z - cw^k)$$

Thus, the sinusoidal spiral equation reformulates as follows:

$$\begin{aligned} |z^n - c^n| = c^n \quad &\iff \quad \prod_{k=1}^{n} |z - cw^k| = c^n \\ &\iff \quad \sqrt[n]{\prod_{k=1}^{n} |z - cw^k|} = c \end{aligned}$$

Thus, for a sinusoidal spiral with positive integer parameter, the geometric mean of the distances to the vertices of a regular polygon must equal the radius of the polygon. □

Regarding now the sinusoidal spirals with $n \in -\mathbb{N}$, these are too part of another remarkable family of plane algebraic curves, constructed as follows:

THEOREM 3.15. *Given points in the plane $c_1, \ldots, c_n \in \mathbb{C}$ and a number $d \in \mathbb{R}$, construct the associated stelloid as being the set of points $z \in \mathbb{C}$ verifying*

$$\frac{1}{n} \sum_{k=1}^{n} \alpha_v (z - c_i) = d$$

with α_v denoting the angle with respect to a direction v. Then the stelloid is an algebraic curve, not depending on v, and at the level of examples we have the sinusoidal spirals with $n \in -\mathbb{N}$, including the $n = -1$ line, $n = -2$ hyperbola, and $n = -3$ Humbert cubic.

PROOF. All this is quite self-explanatory, and we will leave the verification of the various generalities regarding the stelloids, as well as the verification of the relation with the sinusoidal spirals with $n \in -\mathbb{N}$, as an instructive exercise. As a bonus exercise, try understanding the precise relation between stelloids, and polynomial lemniscates. □

So long for plane algebraic curves. Needless to say, all the above is old-style, first class mathematics, having countless applications. For instance when doing classical mechanics or electrodynamics, you will certainly meet polynomial lemniscates and stelloids, when looking at the field lines. Also, the image of any circle passing though 0 by $z \to z^2$ is a cardioid, and the famous Mandelbrot set is organized around such a cardioid.

3e. Exercises

Exercises:

Exercise 3.16.

Exercise 3.17.

Exercise 3.18.

Exercise 3.19.

Exercise 3.20.

Exercise 3.21.

Exercise 3.22.

Exercise 3.23.

Bonus exercise.

CHAPTER 4

Polynomials, roots

4a. Resultant, discriminant

We have seen that many questions lead us into computing roots of polynomials. Let us start with something that we know well, but is always good to remember:

PROPOSITION 4.1. *The solutions of $ax^2+bx+c=0$ with $a,b,c\in\mathbb{C}$ are*

$$x_{1,2}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

with the square root of complex numbers being defined as $\sqrt{re^{it}}=\sqrt{r}e^{it/2}$.

PROOF. We can indeed write our equation in the following way:

$$\begin{aligned}ax^2+bx+c=0 &\iff x^2+\frac{b}{a}x+\frac{c}{a}=0\\ &\iff \left(x+\frac{b}{2a}\right)^2=\frac{b^2-4ac}{4a^2}\\ &\iff x+\frac{b}{2a}=\pm\frac{\sqrt{b^2-4ac}}{2a}\end{aligned}$$

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

In degree 3 and higher, we would first like to understand what the analogue of the discriminant $\Delta=b^2-4ac$ is. In order to discuss this question, let us start with:

THEOREM 4.2. *Given a monic polynomial $P\in\mathbb{C}[X]$, factorized as*

$$P=(X-a_1)\ldots(X-a_k)$$

the following happen:

(1) *The coefficients of P are symmetric functions in $a_1,\ldots,a_k$.*
(2) *The symmetric functions in $a_1,\ldots,a_k$ are polynomials in the coefficients of P.*

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

(1) By expanding our polynomial, we have the following formula:

$$P=\sum_{r=0}^{k}(-1)^r\sum_{i_1<\ldots<i_r}a_{i_1}\ldots a_{i_r}\cdot X^{k-r}$$

Thus the coefficients of P are, up to some signs, the following functions:

$$f_r = \sum_{i_1<\ldots<i_r} a_{i_1} \ldots a_{i_r}$$

But these are indeed symmetric functions in $a_1, \ldots, a_k$, as claimed.

(2) Conversely now, let us look at the symmetric functions in the roots $a_1, \ldots, a_k$. These appear as linear combinations of the basic symmetric functions, given by:

$$S_r = \sum_i a_i^r$$

Moreover, when allowing polynomials instead of linear combinations, we need in fact only the first k such sums, namely $S_1, \ldots, S_k$. That is, the symmetric functions $\mathcal{F}$ in our variables $a_1, \ldots, a_k$, with integer coefficients, appear as follows:

$$\mathcal{F} = \mathbb{Z}[S_1, \ldots, S_k]$$

(3) The point now is that, alternatively, the symmetric functions in our variables $a_1, \ldots, a_k$ appear as well as linear combinations of the functions f_r that we found in (1), and that when allowing polynomials instead of linear combinations, we need in fact only the first k functions, namely $f_1, \ldots, f_k$. That is, we have as well:

$$\mathcal{F} = \mathbb{Z}[f_1, \ldots, f_k]$$

But this gives the result, because we can pass from $\{S_r\}$ to $\{f_r\}$, and vice versa.

(4) This was for the idea, and in practice now up to you to clarify all the details. In fact, we will also need in what follows the extension of all this to the case where P is no longer assumed to be monic, and with this being, again, exercise for you. □

Getting back now to our original question, namely that of deciding whether two polynomials $P, Q \in \mathbb{C}[X]$ have a common root or not, this has the following nice answer:

THEOREM 4.3. *Given two polynomials $P, Q \in \mathbb{C}[X]$, written as*

$$P = c(X - a_1) \ldots (X - a_k) \quad , \quad Q = d(X - b_1) \ldots (X - b_l)$$

the following quantity, which is called resultant of P, Q,

$$R(P, Q) = c^l d^k \prod_{ij} (a_i - b_j)$$

is a certain polynomial in the coefficients of P, Q, with integer coefficients, and we have $R(P, Q) = 0$ precisely when P, Q have a common root.

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

(1) Given two polynomials $P, Q \in \mathbb{C}[X]$, we can certainly construct the quantity $R(P, Q)$ in the statement, with the role of the normalization factor $c^l d^k$ to become clear later on, and then we have $R(P, Q) = 0$ precisely when P, Q have a common root:

$$R(P, Q) = 0 \iff \exists i, j, a_i = b_j$$

(2) As bad news, however, this quantity $R(P, Q)$, defined in this way, is a priori not very useful in practice, because it depends on the roots a_i, b_j of our polynomials P, Q, that we cannot compute in general. However, and here comes our point, as we will prove below, it turns out that $R(P, Q)$ is in fact a polynomial in the coefficients of P, Q, with integer coefficients, and this is where the power of $R(P, Q)$ comes from.

(3) You might perhaps say, nice, but why not doing things the other way around, that is, formulating our theorem with the explicit formula of $R(P, Q)$, in terms of the coefficients of P, Q, and then proving that we have $R(P, Q) = 0$, via roots and everything. Good point, but this is not exactly obvious, the formula of $R(P, Q)$ in terms of the coefficients of P, Q being something terribly complicated. In short, trust me, let us prove our theorem as stated, and for alternative formulae of $R(P, Q)$, we will see later.

(4) Getting started now, let us expand the formula of $R(P, Q)$, by making all the multiplications there, abstractly, in our head. Everything being symmetric in $a_1, \ldots, a_k$, we obtain in this way certain symmetric functions in these variables, which will be therefore certain polynomials in the coefficients of P. Moreover, due to our normalization factor c^l, these polynomials in the coefficients of P will have integer coefficients.

(5) With this done, let us look now what happens with respect to the remaining variables $b_1, \ldots, b_l$, which are the roots of Q. Once again what we have here are certain symmetric functions in these variables $b_1, \ldots, b_l$, and these symmetric functions must be certain polynomials in the coefficients of Q. Moreover, due to our normalization factor d^k, these polynomials in the coefficients of Q will have integer coefficients.

(6) Thus, we are led to the conclusion in the statement, that $R(P, Q)$ is a polynomial in the coefficients of P, Q, with integer coefficients, and with the remark that the $c^l d^k$ factor is there for these latter coefficients to be indeed integers, instead of rationals. □

All the above might seem a bit complicated, so as an illustration, let us work out an example. Consider the case of a polynomial of degree 2, and a polynomial of degree 1:

$$P = ax^2 + bx + c \quad , \quad Q = dx + e$$

In order to compute the resultant, let us factorize our polynomials:

$$P = a(x - p)(x - q) \quad , \quad Q = d(x - r)$$

The resultant can be then computed as follows, by using the method above:

$$\begin{aligned} R(P,Q) &= ad^2(p-r)(q-r)\\ &= ad^2(pq-(p+q)r+r^2)\\ &= cd^2+bd^2r+ad^2r^2\\ &= cd^2-bde+ae^2 \end{aligned}$$

Finally, observe that $R(P,Q)=0$ corresponds indeed to the fact that P,Q have a common root. Indeed, the root of Q is $r=-e/d$, and we have:

$$P(r)=\frac{ae^2}{d^2}-\frac{be}{d}+c=\frac{R(P,Q)}{d^2}$$

Regarding now the explicit formula of the resultant $R(P,Q)$, this is something quite complicated, and there are several methods for dealing with this problem. We have:

THEOREM 4.4. *The resultant of two polynomials, written as*

$$P=p_kX^k+\ldots+p_1X+p_0\quad,\quad Q=q_lX^l+\ldots+q_1X+q_0$$

appears as the determinant of an associated matrix, as follows,

$$R(P,Q)=\begin{vmatrix} p_k & & & q_l & & \\ \vdots & \ddots & & \vdots & \ddots & \\ p_0 & & p_k & q_0 & & q_l \\ & \ddots & \vdots & & \ddots & \vdots \\ & & p_0 & & & q_0 \end{vmatrix}$$

with the matrix having size $k+l$, and having 0 coefficients at the blank spaces.

PROOF. This is something clever, due to Sylvester, as follows:

(1) Consider the vector space $\mathbb{C}_k[X]$ formed by the polynomials of degree $<k$:

$$\mathbb{C}_k[X]=\left\{P\in\mathbb{C}[X]\,\middle|\,\deg P<k\right\}$$

This is a vector space of dimension k, having as basis the monomials $1,X,\ldots,X^{k-1}$. Now given polynomials P,Q as in the statement, consider the following linear map:

$$\Phi:\mathbb{C}_l[X]\times\mathbb{C}_k[X]\to\mathbb{C}_{k+l}[X]\quad,\quad(A,B)\to AP+BQ$$

(2) Our first claim is that with respect to the standard bases for all the vector spaces involved, namely those consisting of the monomials $1,X,X^2,\ldots$, the matrix of Φ is the matrix in the statement. But this is something which is clear from definitions.

(3) Our second claim is that $\det\Phi=0$ happens precisely when P,Q have a common root. Indeed, our polynomials P,Q having a common root means that we can find A,B such that $AP+BQ=0$, and so that $(A,B)\in\ker\Phi$, which reads $\det\Phi=0$.

(4) Finally, our claim is that we have $\det\Phi = R(P,Q)$. But this follows from the uniqueness of the resultant, up to a scalar, and with this uniqueness property being elementary to establish, along the lines of the proofs of Theorems 4.2 and 4.3. □

In what follows we will not really need the above formula, so let us just check now that this formula works indeed. Consider our favorite polynomials, as before:

$$P = ax^2 + bx + c \quad , \quad Q = dx + e$$

According to the above result, the resultant should be then, as it should:

$$R(P,Q) = \begin{vmatrix} a & d & 0 \\ b & e & d \\ c & 0 & e \end{vmatrix} = ae^2 - bde + cd^2$$

We can go back now to our original question, and we have:

THEOREM 4.5. *Given a polynomial $P \in \mathbb{C}[X]$, written as*

$$P(X) = aX^N + bX^{N-1} + cX^{N-2} + \ldots$$

its discriminant, defined as being the following quantity,

$$\Delta(P) = \frac{(-1)^{\binom{N}{2}}}{a} R(P,P')$$

is a polynomial in the coefficients of P, with integer coefficients, and $\Delta(P) = 0$ happens precisely when P has a double root.

PROOF. The fact that the discriminant $\Delta(P)$ is a polynomial in the coefficients of P, with integer coefficients, comes from Theorem 4.3, coupled with the fact that the division by the leading coefficient a is indeed possible, under $\mathbb{Z}$, as being shown by the following formula, which is written of course a bit informally, coming from Theorem 4.4:

$$R(P,P') = \begin{vmatrix} a & & & Na & & \\ \vdots & \ddots & & \vdots & \ddots & \\ z & & a & y & & Na \\ & \ddots & \vdots & & \ddots & \vdots \\ & & z & & & y \end{vmatrix}$$

Also, the fact that we have $\Delta(P) = 0$ precisely when P has a double root is clear from Theorem 4.3. Finally, let us mention that the sign $(-1)^{\binom{N}{2}}$ is there for various reasons, including the compatibility with some well-known formulae, at small values of $N \in \mathbb{N}$, such as $\Delta(P) = b^2 - 4ac$ in degree 2, that we will discuss in a moment. □

As already mentioned, by using Theorem 4.4, we have an explicit formula for the discriminant, as the determinant of a certain matrix. There is a lot of theory here, and in order to get into this, let us first see what happens in degree 2. Here we have:

$$P = aX^2 + bX + c \quad , \quad P' = 2aX + b$$

Thus, the resultant is given by the following formula:

$$\begin{aligned} R(P, P') &= ab^2 - b(2a)b + c(2a)^2 \\ &= 4a^2c - ab^2 \\ &= -a(b^2 - 4ac) \end{aligned}$$

It follows that the discriminant of our polynomial is, as it should:

$$\Delta(P) = b^2 - 4ac$$

Alternatively, we can use the formula in Theorem 3.4, and we obtain:

$$\begin{aligned} \Delta(P) = &= -\frac{1}{a}\begin{vmatrix} a & 2a & \\ b & b & 2a \\ c & & b \end{vmatrix} \\ &= -\begin{vmatrix} 1 & 2 & \\ b & b & 2a \\ c & & b \end{vmatrix} \\ &= -b^2 + 2(b^2 - 2ac) \\ &= b^2 - 4ac \end{aligned}$$

We will be back later to such formulae, in degree 3, and in degree 4 as well, with the comment however, coming in advance, that these formulae are not very beautiful.

At the theoretical level now, we have the following result, which is not trivial:

THEOREM 4.6. *The discriminant of a polynomial P is given by the formula*

$$\Delta(P) = a^{2N-2}\prod_{i<j}(r_i - r_j)^2$$

where a is the leading coefficient, and $r_1, \ldots, r_N$ are the roots.

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

(1) The first thought goes to the formula in Theorem 4.3, so let us see what that formula teaches us, in the case $Q = P'$. Let us write P, P' as follows:

$$P = a(x - r_1)\ldots(x - r_N)$$

$$P' = Na(x - p_1)\ldots(x - p_{N-1})$$

According to Theorem 4.3, the resultant of P, P' is then given by:

$$R(P,P') = a^{N-1}(Na)^N \prod_{ij}(r_i - p_j)$$

And bad news, this is not exactly what we wished for, namely the formula in the statement. That is, we are on the good way, but certainly have to work some more.

(2) Obviously, we must get rid of the roots $p_1, \ldots, p_{N-1}$ of the polynomial P'. In order to do this, let us rewrite the formula that we found in (1) in the following way:

$$\begin{aligned} R(P,P') &= N^N a^{2N-1} \prod_i \left(\prod_j (r_i - p_j) \right) \\ &= N^N a^{2N-1} \prod_i \frac{P'(r_i)}{Na} \\ &= a^{N-1} \prod_i P'(r_i) \end{aligned}$$

(3) In order to compute now P', and more specifically the values $P'(r_i)$ that we are interested in, we can use the Leibnitz rule. So, consider our polynomial:

$$P(x) = a(x - r_1) \ldots (x - r_N)$$

The Leibnitz rule for derivatives tells us that $(fg)' = f'g + fg'$, but then also that $(fgh)' = f'gh + fg'h + fgh'$, and so on. Thus, for our polynomial, we obtain:

$$P'(x) = a \sum_i (x - r_1) \ldots \underbrace{(x - r_i)}_{missing} \ldots (x - r_N)$$

Now when applying this formula to one of the roots r_i, we obtain:

$$P'(r_i) = a(r_i - r_1) \ldots \underbrace{(r_i - r_i)}_{missing} \ldots (r_i - r_N)$$

By making now the product over all indices i, this gives the following formula:

$$\prod_i P'(r_i) = a^N \prod_{i \neq j} (r_i - r_j)$$

(4) Time now to put everything together. By taking the formula in (2), making the normalizations in Theorem 3.5, and then using the formula found in (3), we obtain:

$$\begin{aligned} \Delta(P) &= (-1)^{\binom{N}{2}} a^{N-2} \prod_i P'(r_i) \\ &= (-1)^{\binom{N}{2}} a^{2N-2} \prod_{i \neq j} (r_i - r_j) \end{aligned}$$

(5) This is already a nice formula, which is very useful in practice, and that we can safely keep as a conclusion, to our computations. However, we can do slightly better, by grouping opposite terms. Indeed, this gives the following formula:

$$\begin{aligned}\Delta(P) &= (-1)^{\binom{N}{2}}a^{2N-2}\prod_{i\neq j}(r_i-r_j)\\ &= (-1)^{\binom{N}{2}}a^{2N-2}\prod_{i<j}(r_i-r_j)\cdot\prod_{i>j}(r_i-r_j)\\ &= (-1)^{\binom{N}{2}}a^{2N-2}\prod_{i<j}(r_i-r_j)\cdot(-1)^{\binom{N}{2}}\prod_{i<j}(r_i-r_j)\\ &= a^{2N-2}\prod_{i<j}(r_i-r_j)^2\end{aligned}$$

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

As applications now, the formula in Theorem 4.6 is quite useful for the real polynomials $P\in\mathbb{R}[X]$ in small degree, because it allows to say when the roots are real, or complex, or at least have some partial information about this. For instance, we have:

PROPOSITION 4.7. *Consider a polynomial with real coefficients, $P\in\mathbb{R}[X]$, assumed for simplicity to have nonzero discriminant, $\Delta\neq 0$.*

(1) *In degree 2, the roots are real when $\Delta>0$, and complex when $\Delta<0$.*
(2) *In degree 3, all roots are real precisely when $\Delta>0$.*

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

(1) The first assertion is something that we certainly know, coming from Proposition 4.1, but let us see how this comes via the formula in Theorem 4.6, namely:

$$\Delta(P)=a^{2N-2}\prod_{i<j}(r_i-r_j)^2$$

In degree $N=2$, this formula looks as follows, with r_1,r_2 being the roots:

$$\Delta(P)=a^2(r_1-r_2)^2$$

Thus $\Delta>0$ amounts in saying that we have $(r_1-r_2)^2>0$. Now since r_1,r_2 are conjugate, and with this being something trivial, meaning no need here for the computations in Proposition 4.1, we conclude that $\Delta>0$ means that r_1,r_2 are real, as stated.

(2) In degree $N=3$ now, we know from analysis that P has at least one real root, and the problem is whether the remaining 2 roots are real, or complex conjugate. For this purpose, we can use the formula in Theorem 4.6, which in degree 3 reads:

$$\Delta(P)=a^4(r_1-r_2)^2(r_1-r_3)^2(r_2-r_3)^2$$

We can see that in the case $r_1, r_2, r_3 \in \mathbb{R}$, we have $\Delta(P) > 0$. Conversely now, assume that $r_1 = r$ is the real root, coming from analysis, and that the other roots are $r_2 = z$ and $r_3 = \bar{z}$, with z being a complex number, which is not real. We have then:

$$\begin{aligned}\Delta(P) &= a^4(r-z)^2(r-\bar{z})^2(z-\bar{z})^2\\ &= a^4|r-z|^4(2iIm(z))^2\\ &= -4a^4|r-z|^4Im(z)^2\\ &< 0\end{aligned}$$

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

In relation with the above, for our result to be truly useful, we must of course compute the discriminant in degree 3. We will do this in the next section.

Finally, as another application of all this, worth mentioning, we have:

THEOREM 4.8. *The diagonalizable matrices are dense.*

PROOF. As a first observation, this is something extremely useful, more or less allowing you in practice to assume that any matrix $A \in M_N(\mathbb{C})$ is diagonalizable, but of course do not try this at home, unless you know what you're doing. As for the proof, this is non-trivial, and there are actually two standard proofs, both non-trivial, as follows:

(1) Via the pedestrian way, by using the Jordan form. Here you have to learn well the Jordan form, and good luck with that, and once that done, you can argue that by perturbing the Jordan blocks, in the obvious way, you can arrange up to epsilon as for your matrix to have distinct eigenvalues, and so to be diagonalizable.

(2) As a geometry king, using the discriminant. Indeed, for a matrix $A \in M_N(\mathbb{C})$, with characteristic polynomial P_A, having distinct eigenvalues means:

$$\Delta(P_A) \neq 0$$

But this is the complement of a hypersurface, which is dense, and since all these matrices are diagonalizable, the diagonalizable matrices are dense too. Just like that. □

4b. Cardano formula

Let us work out now what happens in degree 3. Here the result is as follows:

THEOREM 4.9. *The discriminant of a degree* 3 *polynomial,*

$$P = aX^3 + bX^2 + cX + d$$

is the number $\Delta(P) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$.

PROOF. We have two methods available, based on Theorem 4.3 and Theorem 4.4, and both being instructive, we will try them both. The computations are as follows:

(1) Let us first go the pedestrian way, based on the definition of the resultant, from Theorem 4.3. Consider two polynomials, of degree 3 and degree 2, written as follows:

$$P = aX^3 + bX^2 + cX + d$$

$$Q = eX^2 + fX + g = e(X - s)(X - t)$$

The resultant of these two polynomials is then given by:

$$\begin{aligned} R(P,Q) &= a^2e^3(p-s)(p-t)(q-s)(q-t)(r-s)(r-t) \\ &= a^2 \cdot e(p-s)(p-t) \cdot e(q-s)(q-t) \cdot e(r-s)(r-t) \\ &= a^2Q(p)Q(q)Q(r) \\ &= a^2(ep^2+fp+g)(eq^2+fq+g)(er^2+fr+g) \end{aligned}$$

By expanding, we obtain the following formula for this resultant:

$$\begin{aligned} \frac{R(P,Q)}{a^2} &= e^3p^2q^2r^2 + e^2f(p^2q^2r + p^2qr^2 + pq^2r^2) \\ &+ e^2g(p^2q^2 + p^2r^2 + q^2r^2) + ef^2(p^2qr + pq^2r + pqr^2) \\ &+ efg(p^2q + pq^2 + p^2r + pr^2 + q^2r + qr^2) + f^3pqr \\ &+ eg^2(p^2 + q^2 + r^2) + f^2g(pq + pr + qr) \\ &+ fg^2(p + q + r) + g^3 \end{aligned}$$

Note in passing that we have 27 terms on the right, as we should, and with this kind of check being mandatory, when doing such computations. Next, we have:

$$p + q + r = -\frac{b}{a} \quad , \quad pq + pr + qr = \frac{c}{a} \quad , \quad pqr = -\frac{d}{a}$$

By using these formulae, we can produce some more, as follows:

$$p^2 + q^2 + r^2 = (p + q + r)^2 - 2(pq + pr + qr) = \frac{b^2}{a^2} - \frac{2c}{a}$$

$$p^2q + pq^2 + p^2r + pr^2 + q^2r + qr^2 = (p + q + r)(pq + pr + qr) - 3pqr = -\frac{bc}{a^2} + \frac{3d}{a}$$

$$p^2q^2 + p^2r^2 + q^2r^2 = (pq + pr + qr)^2 - 2pqr(p + q + r) = \frac{c^2}{a^2} - \frac{2bd}{a^2}$$

By plugging now this data into the formula of $R(P,Q)$, we obtain:

$$\begin{aligned}R(P,Q) &= a^2e^3\cdot\frac{d^2}{a^2}-a^2e^2f\cdot\frac{cd}{a^2}+a^2e^2g\left(\frac{c^2}{a^2}-\frac{2bd}{a^2}\right)+a^2ef^2\cdot\frac{bd}{a^2}\\ &+ a^2efg\left(-\frac{bc}{a^2}+\frac{3d}{a}\right)-a^2f^3\cdot\frac{d}{a}\\ &+ a^2eg^2\left(\frac{b^2}{a^2}-\frac{2c}{a}\right)+a^2f^2g\cdot\frac{c}{a}-a^2fg^2\cdot\frac{b}{a}+a^2g^3\end{aligned}$$

Thus, we have the following formula for the resultant:

$$\begin{aligned}R(P,Q) &= d^2e^3-cde^2f+c^2e^2g-2bde^2g+bdef^2-bcefg+3adefg\\ &- adf^3+b^2eg^2-2aceg^2+acf^2g-abfg^2+a^2g^3\end{aligned}$$

Getting back now to our discriminant problem, with $Q=P'$, which corresponds to $e=3a$, $f=2b$, $g=c$, we obtain the following formula:

$$\begin{aligned}R(P,P') &= 27a^3d^2-18a^2bcd+9a^2c^3-18a^2bcd+12ab^3d-6ab^2c^2+18a^2bcd\\ &- 8ab^3d+3ab^2c^2-6a^2c^3+4ab^2c^2-2ab^2c^2+a^2c^3\end{aligned}$$

By simplifying terms, and dividing by a, we obtain the following formula:

$$-\Delta(P)=27a^2d^2-18abcd+4ac^3+4b^3d-b^2c^2$$

But this gives the formula in the statement, namely:

$$\Delta(P)=b^2c^2-4ac^3-4b^3d-27a^2d^2+18abcd$$

(2) Let us see as well how the computation does, by using Theorem 4.4, which is our most advanced tool, so far. Consider a polynomial of degree 3, and its derivative:

$$P=aX^3+bX^2+cX+d$$

$$P'=3aX^2+2bX+c$$

By using now Theorem 4.4 and computing the determinant, we obtain:

$$\begin{aligned}
R(P,P') &= \begin{vmatrix} a & & 3a & & \\ b & a & 2b & 3a & \\ c & b & c & 2b & 3a \\ d & c & & c & 2b \\ & d & & & c \end{vmatrix} \\
&= \begin{vmatrix} a & & & & \\ b & a & -b & 3a & \\ c & b & -2c & 2b & 3a \\ d & c & -3d & c & 2b \\ & d & & & c \end{vmatrix} \\
&= a\begin{vmatrix} a & -b & 3a & \\ b & -2c & 2b & 3a \\ c & -3d & c & 2b \\ d & & & c \end{vmatrix} \\
&= -ad\begin{vmatrix} -b & 3a & \\ -2c & 2b & 3a \\ -3d & c & 2b \end{vmatrix} + ac\begin{vmatrix} a & -b & 3a \\ b & -2c & 2b \\ c & -3d & c \end{vmatrix} \\
&= -ad(-4b^3 - 27a^2d + 12abc + 3abc) \\
&\quad +ac(-2ac^2 - 2b^2c - 9abd + 6ac^2 + b^2c + 6abd) \\
&= a(4b^3d + 27a^2d^2 - 15abcd + 4ac^3 - b^2c^2 - 3abcd) \\
&= a(4b^3d + 27a^2d^2 - 18abcd + 4ac^3 - b^2c^2)
\end{aligned}$$

Now according to Theorem 4.5, the discriminant of our polynomial is given by:

$$\begin{aligned}
\Delta(P) &= -\frac{R(P,P')}{a} \\
&= -4b^3d - 27a^2d^2 + 18abcd - 4ac^3 + b^2c^2 \\
&= b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd
\end{aligned}$$

Thus, we have again obtained the formula in the statement. □

Still talking degree 3 equations, let us try now to solve such an equation $P = 0$, with $P = aX^3 + bX^2 + cX + d$ as above. By linear transformations we can assume $a = 1, b = 0$, and then it is convenient to write $c = 3p, d = 2q$. Thus, our equation becomes:

$$x^3 + 3px + 2q = 0$$

Regarding such equations, many things can be said, and to start with, we have the following famous result, dealing with real roots, due to Cardano:

THEOREM 4.10. *For a normalized degree 3 equation, namely*

$$x^3+3px+2q=0$$

the discriminant is $\Delta=-108(p^3+q^2)$. Assuming $p,q\in\mathbb{R}$ and $\Delta<0$, the number

$$x=\sqrt[3]{-q+\sqrt{p^3+q^2}}+\sqrt[3]{-q-\sqrt{p^3+q^2}}$$

is a real solution of our equation.

PROOF. The formula of Δ is clear from definitions, and with $108=4\times 27$. Now with x as in the statement, by using $(a+b)^3=a^3+b^3+3ab(a+b)$, we have:

$$\begin{aligned}x^3&=\left(\sqrt[3]{-q+\sqrt{p^3+q^2}}+\sqrt[3]{-q-\sqrt{p^3+q^2}}\right)^3\\&=-2q+3\sqrt[3]{-q+\sqrt{p^3+q^2}}\cdot\sqrt[3]{-q-\sqrt{p^3+q^2}}\cdot x\\&=-2q+3\sqrt[3]{q^2-p^3-q^2}\cdot x\\&=-2q-3px\end{aligned}$$

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

Regarding the other roots, we know from Proposition 4.7 that these are both real when $\Delta<0$, and complex conjugate when $\Delta<0$. Thus, in the context of Theorem 4.10, the other two roots are complex conjugate, the formula for them being as follows:

PROPOSITION 4.11. *For a normalized degree 3 equation, namely*

$$x^3+3px+2q=0$$

with $p,q\in\mathbb{R}$ and discriminant $\Delta=-108(p^3+q^2)$ negative, $\Delta<0$, the numbers

$$z=w\sqrt[3]{-q+\sqrt{p^3+q^2}}+w^2\sqrt[3]{-q-\sqrt{p^3+q^2}}$$

$$\bar{z}=w^2\sqrt[3]{-q+\sqrt{p^3+q^2}}+w\sqrt[3]{-q-\sqrt{p^3+q^2}}$$

with $w=e^{2\pi i/3}$ are the complex conjugate solutions of our equation.

PROOF. As before, by using $(a+b)^3=a^3+b^3+3ab(a+b)$, we have:

$$\begin{aligned}z^3&=\left(w\sqrt[3]{-q+\sqrt{p^3+q^2}}+w^2\sqrt[3]{-q-\sqrt{p^3+q^2}}\right)^3\\&=-2q+3\sqrt[3]{-q+\sqrt{p^3+q^2}}\cdot\sqrt[3]{-q-\sqrt{p^3+q^2}}\cdot z\\&=-2q+3\sqrt[3]{q^2-p^3-q^2}\cdot z\\&=-2q-3pz\end{aligned}$$

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

As a conclusion, we have the following statement, unifying the above:

THEOREM 4.12. *For a normalized degree* 3 *equation, namely*

$$x^3+3px+2q=0$$

the discriminant is $\Delta=-108(p^3+q^2)$. *Assuming* $p,q\in\mathbb{R}$ *and* $\Delta<0$, *the numbers*

$$x=w\sqrt[3]{-q+\sqrt{p^3+q^2}}+w^2\sqrt[3]{-q-\sqrt{p^3+q^2}}$$

with $w=1,e^{2\pi i/3},e^{4\pi i/3}$ *are the solutions of our equation.*

PROOF. This follows indeed from Theorem 4.10 and Proposition 4.11. Alternatively, we can redo the computation in their proof, which was nearly identical anyway, in the present setting, with x being given by the above formula, by using $w^3=1$. □

As a comment here, the formula in Theorem 4.12 holds of course in the case $\Delta>0$ too, and also when the coefficients are complex numbers, $p,q\in\mathbb{C}$, and this due to the fact that the proof rests on the nearly trivial computation from the proof of Theorem 4.10, or of Proposition 4.11. However, these extensions are quite often not very useful, because when it comes to extract all the above square and cubic roots, for complex numbers, you can well end up with the initial question, the one that you started with.

Thus, as a conclusion to this, Theorem 4.12 as formulated above is what can be best said about the degree 3 equations. There are of course many versions of it, and slight generalizations, but in practice, Theorem 4.12 is what mostly matters.

4c. Higher degree

In higher degree things become quite complicated. In degree 4, to start with, we first have the following result, dealing with the discriminant and its applications:

THEOREM 4.13. *The discriminant of* $P=ax^4+bx^3+cx^2+dx+e$ *is given by the following formula:*

$$\begin{aligned}\Delta &= 256a^3e^3-192a^2bde^2-128a^2c^2e^2+144a^2cd^2e-27a^2d^4\\ &\quad +144ab^2ce^2-6ab^2d^2e-80abc^2de+18abcd^3+16ac^4e\\ &\quad -4ac^3d^2-27b^4e^2+18b^3cde-4b^3d^3-4b^2c^3e+b^2c^2d^2\end{aligned}$$

In the case $\Delta<0$ *we have* 2 *real roots and* 2 *complex conjugate roots, and in the case* $\Delta>0$ *the roots are either all real or all complex.*

PROOF. The formula of Δ follows from the definition of the discriminant, from Theorem 4.5, with the resultant computed via Theorem 4.4, as follows:

$$\Delta=\frac{1}{a}\begin{vmatrix}a & & & 4a & & & \\ b & a & & 3b & 4a & & \\ c & b & a & 2c & 3b & 4a & \\ d & c & b & d & 2c & 3b & 4a\\ e & d & c & & d & 2c & 3b\\ & e & d & & & d & 2c\\ & & e & & & & d\end{vmatrix}$$

As for the last assertion, the study here is routine, a bit as in degree 3. □

In practice, as in degree 3, we can do first some manipulations on our polynomials, as to have them in simpler form, and we have the following version of Theorem 4.13:

PROPOSITION 4.14. *The discriminant of $P=x^4+cx^2+dx+e$, normalized degree 4 polynomial, is given by the following formula:*

$$\Delta=16c^4e-4c^3d^2-128c^2e^2+144cd^2e-27d^4+256e^3$$

As before, if $\Delta<0$ we have 2 real roots and 2 complex conjugate roots, and if $\Delta>0$ the roots are either all real or all complex.

PROOF. This is a consequence of Theorem 4.13, with $a=1,b=0$, but we can deduce this as well directly. Indeed, the formula of Δ follows, quite easily, from:

$$\Delta=\begin{vmatrix}1 & & & 4 & & & \\ & 1 & & & 4 & & \\ c & & 1 & 2c & & 4 & \\ d & c & & d & 2c & & 4\\ e & d & c & & d & 2c & \\ & e & d & & & d & 2c\\ & & e & & & & d\end{vmatrix}$$

As for the last assertion, this is something that we know, from Theorem 4.13. □

We still have some work to do. Indeed, looking back at what we did in degree 3, the passage there from Theorem 4.9 to Theorem 4.10 was made of two operations, namely "depressing" the equation, that is, getting rid of the next-to-highest term, and then rescaling the coefficients, as for the formula of Δ to become as simple as possible.

In our present setting now, degree 4, with the depressing done as above, in Proposition 4.14, it remains to rescale the coefficients, as for the formula of Δ to become as simple as possible. And here, a bit of formula hunting, in relation with $2,3$ powers, leads to:

THEOREM 4.15. *The discriminant of a normalized degree* 4 *polynomial, written as*

$$P = x^4 + 6px^2 + 4qx + 3r$$

is given by the following formula:

$$\Delta = 256 \times 27 \times \left(9p^4r - 2p^3q^2 - 6p^2r^2 + 6pq^2r - q^4 + r^3\right)$$

In the case $\Delta < 0$ *we have* 2 *real roots and* 2 *complex conjugate roots, and in the case* $\Delta > 0$ *the roots are either all real or all complex.*

PROOF. This follows from Proposition 4.14, with $c = 6p, d = 4q, e = 3r$, but we can deduce this as well directly. Indeed, the formula of Δ follows, quite easily, from:

$$\Delta = \begin{vmatrix} 1 & & & 4 & & & \\ & 1 & & & 4 & & \\ 6p & & 1 & 12p & & 4 & \\ 4q & 6p & & 4q & 12p & & 4 \\ 3r & 4q & 6p & & 4q & 12p & \\ & 3r & 4q & & & 4q & 12p \\ & & 3r & & & & 4q \end{vmatrix}$$

As for the last assertion, this is something that we know from Theorem 4.13. □

Time now to get to the real thing, solving the equation. We have here:

THEOREM 4.16. *The roots of a normalized degree* 4 *equation, written as*

$$x^4 + 6px^2 + 4qx + 3r = 0$$

are as follows, with y *satisfying the equation* $(y^2 - 3r)(y - 3p) = 2q^2$*,*

$$x_1 = \frac{1}{\sqrt{2}}\left(-\sqrt{y - 3p} + \sqrt{-y - 3p + \frac{4q}{\sqrt{2y - 6p}}}\right)$$

$$x_2 = \frac{1}{\sqrt{2}}\left(-\sqrt{y - 3p} - \sqrt{-y - 3p + \frac{4q}{\sqrt{2y - 6p}}}\right)$$

$$x_3 = \frac{1}{\sqrt{2}}\left(\sqrt{y - 3p} + \sqrt{-y - 3p - \frac{4q}{\sqrt{2y - 6p}}}\right)$$

$$x_4 = \frac{1}{\sqrt{2}}\left(\sqrt{y - 3p} - \sqrt{-y - 3p - \frac{4q}{\sqrt{2y - 6p}}}\right)$$

and with y *being computable via the Cardano formula.*

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

(1) To start with, let us write our equation in the following form:

$$x^4 = -6px^2 - 4qx - 3r$$

The idea will be that of adding a suitable common term, to both sides, as to make square on both sides, as to eventually end with a sort of double quadratic equation. For this purpose, our claim is that what we need is a number y satisfying:

$$(y^2 - 3r)(y - 3p) = 2q^2$$

Indeed, assuming that we have this number y, our equation becomes:

$$\begin{aligned}(x^2 + y)^2 &= x^4 + 2x^2y + y^2\\ &= -6px^2 - 4qx - 3r + 2x^2y + y^2\\ &= (2y - 6p)x^2 - 4qx + y^2 - 3r\\ &= (2y - 6p)x^2 - 4qx + \frac{2q^2}{y - 3p}\\ &= \left(\sqrt{2y - 6p}\cdot x - \frac{2q}{\sqrt{2y - 6p}}\right)^2\end{aligned}$$

(2) Which looks very good, leading us to the following degree 2 equations:

$$x^2 + y + \sqrt{2y - 6p}\cdot x - \frac{2q}{\sqrt{2y - 6p}} = 0$$

$$x^2 + y - \sqrt{2y - 6p}\cdot x + \frac{2q}{\sqrt{2y - 6p}} = 0$$

Now let us write these two degree 2 equations in standard form, as follows:

$$x^2 + \sqrt{2y - 6p}\cdot x + \left(y - \frac{2q}{\sqrt{2y - 6p}}\right) = 0$$

$$x^2 - \sqrt{2y - 6p}\cdot x + \left(y + \frac{2q}{\sqrt{2y - 6p}}\right) = 0$$

(3) Regarding the first equation, the solutions there are as follows:

$$x_1 = \frac{1}{2}\left(-\sqrt{2y - 6p} + \sqrt{-2y - 6p + \frac{8q}{\sqrt{2y - 6p}}}\right)$$

$$x_2 = \frac{1}{2}\left(-\sqrt{2y - 6p} - \sqrt{-2y - 6p + \frac{8q}{\sqrt{2y - 6p}}}\right)$$

As for the second equation, the solutions there are as follows:

$$x_3=\frac{1}{2}\left(\sqrt{2y-6p}+\sqrt{-2y-6p-\frac{8q}{\sqrt{2y-6p}}}\right)$$

$$x_4=\frac{1}{2}\left(\sqrt{2y-6p}-\sqrt{-2y-6p-\frac{8q}{\sqrt{2y-6p}}}\right)$$

(4) Now by cutting a $\sqrt{2}$ factor from everything, this gives the formulae in the statement. As for the last claim, regarding the nature of y, this comes from Cardano. □

We still have to compute the number y appearing in the above via Cardano, and the result here, adding to what we already have in Theorem 4.16, is as follows:

THEOREM 4.17 (continuation). *The value of y in the previous theorem is*

$$y=t+p+\frac{a}{t}$$

where the number t is given by the formula

$$t=\sqrt[3]{b+\sqrt{b^2-a^3}}$$

with $a=p^2+r$ and $b=2p^2-3pr+q^2$.

PROOF. The legend goes that this is what comes from Cardano, but depressing and normalizing and solving $(y^2-3r)(y-3p)=2q^2$ makes it for too many operations, so the most pragmatic is to simply check this equation. With y as above, we have:

$$\begin{aligned}y^2-3r&=t^2+2pt+(p^2+2a)+\frac{2pa}{t}+\frac{a^2}{t^2}-3r\\&=t^2+2pt+(3p^2-r)+\frac{2pa}{t}+\frac{a^2}{t^2}\end{aligned}$$

With this in hand, we have the following computation:

$$\begin{aligned}(y^2-3r)(y-3p)&=\left(t^2+2pt+(3p^2-r)+\frac{2pa}{t}+\frac{a^2}{t^2}\right)\left(t-2p+\frac{a}{t}\right)\\&=t^3+(a-4p^2+3p^2-r)t+(2pa-6p^3+2pr+2pa)\\&\quad+(3p^2a-ra-4p^2a+a^2)\frac{1}{t}+\frac{a^3}{t^3}\\&=t^3+(a-p^2-r)t+2p(2a-3p^2+r)+a(a-p^2-r)\frac{1}{t}+\frac{a^3}{t^3}\\&=t^3+2p(-p^2+3r)+\frac{a^3}{t^3}\end{aligned}$$

Now by using the formula of t in the statement, this gives:

$$\begin{aligned}(y^2-3r)(y-3p) &= b+\sqrt{b^2-a^3}-4p^2+6pr+\frac{a^3}{b+\sqrt{b^2-a^3}}\\ &= b+\sqrt{b^2-a^3}-4p^2+6pr+b-\sqrt{b^2-a^3}\\ &= 2b-4p^2+6pr\\ &= 2(2p^2-3pr+q^2)-4p^2+6pr\\ &= 2q^2\end{aligned}$$

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

In degree 5 and more, things become complicated. However, we have some arithmetic tricks here, for computing the integer or rational roots of polynomials having integer of rational coefficients. There are a lot of analytic tricks too, both real and complex.

4d. Galois theory

We discuss now Galois theory, and its applications to degree 5 equations. Let us start with a basic result regarding the arbitrary fields F and their structure, as follows:

THEOREM 4.18. *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 number p must be prime, we have a field embedding $\mathbb{F}_p\subset F$, and $q=|F|$ must be of the form $q=p^k$, with $k\in\mathbb{N}$. Also, we have the formulae

$$(a+b)^p=a^p+b^p\quad,\quad a^q=a$$

valid for any $a,b\in F$, and the Fermat poynomial X^q-X factorizes as:

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

Also, regardless of p, any finite multiplicative subgroup $G\subset F-\{0\}$ must be cyclic.

PROOF. This is a very crowded statement, 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.

(5) The baby Fermat formula $(a + b)^p = a^p + b^p$, which reminds the Fermat little theorem, $a^p = a(p)$ over $\mathbb{Z}$, follows in the same way, namely from the binomial formula, because all the non-trivial binomial coefficients $\binom{p}{s}$ are multiples of p:

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

(6) As for the Fermat formula $a^q = a$ itself, which implies the assertion about $X^q - X$, this follows from the last assertion, which can be proved via some basic arithmetic inside F, and which for $G = F - \{0\}$ itself, with $|F| = q$, gives $a^{q-1} = 1$, for any $a \neq 0$.

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

(8) In order to prove this claim, let $d = (m, n)$, write $d = am + bn$ with $a, b \in \mathbb{Z}$, and set $k = g^a h^b$. 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 $ord(k) = [m, n]$.

(9) 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 (8) 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.

(10) Summarizing, we have proved our claim in (8). 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 (7), namely that the order $m = ord(h)$ of any $h \in G$ satisfies $m|n$.

(11) 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 $g \in G$.

(12) Finally, assuming $|F| = q < \infty$, we know that 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, this gives the Fermat formula $a^q = a$, with of course the remark that this formula trivially holds as well for $a = 0$. □

The above result raises many questions. Since most of these questions seem to have something to do with field extensions, let us start by discussing this. We first have:

THEOREM 4.19. *Given a field extension $E \subset F$, we can talk about its Galois group G, as the group of automorphisms of F fixing E. The intermediate fields*

$$E \subset K \subset F$$

are then in correspondence with the subgroups $H \subset G$, with such a field K corresponding to the subgroup H consisting of automorphisms $g \in G$ fixing K.

PROOF. This is something self-explanatory, and follows indeed from some algebra, under suitable assumptions, in order for that algebra to properly apply. □

Getting now towards polynomials and their roots, we have here:

THEOREM 4.20. *Given a field F and a polynomial $P \in F[X]$, we can talk about the abstract splitting field of P, where this polynomial decomposes as:*

$$P(X) = c\prod_i (X - a_i)$$

In particular, any field F has a certain algebraic closure $\bar{F}$, where all the polynomials $P \in F[X]$, and in fact all polynomials $P \in \bar{F}[X]$ too, have roots.

PROOF. This is again something self-explanatory, which follows from Theorem 4.19 and from some extra algebra, under suitable assumptions, in order for that extra algebra to properly apply. Regarding the construction at the end, as main example here we have $\bar{\mathbb{R}} = \mathbb{C}$. However, as an interesting fact, $\bar{\mathbb{Q}} \subset \mathbb{C}$ is a proper subfield. □

Good news, with this in hand, we can now elucidate the structure of finite fields:

THEOREM 4.21. *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 $\mathbb{F}_p$. In general, this is the splitting field of:*

$$P = X^q - X$$

Moreover, we can construct an explicit model for $\mathbb{F}_q$, at $q = p^2$ or higher, as

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

with $Q \in \mathbb{F}_p[X]$ being a suitable irreducible polynomial, of degree k.

PROOF. There are several assertions here, the idea being as follows:

(1) The first assertion, regarding the existence and uniqueness of $\mathbb{F}_q$, follows from Theorem 4.18 and Theorem 4.20. Indeed, we know from Theorem 4.18 that given a finite field, $|F| = q$ with $k \in \mathbb{N}$, the Fermat polynomial $P = X^q - X$ factorizes as follows:

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

But this shows, via the general theory from Theorem 4.20, that our field F must be the splitting field of P, and so is unique. As for the existence, this follows again from Theorem 4.20, telling us that the splitting field always exists.

(2) In what regards now the modeling of $\mathbb{F}_q$, at $q = p$ there is nothing to do, because we have our usual $\mathbb{F}_p$ here. At $q = p^2$ and higher, we know from commutative algebra that we have an isomorphism as follows, whenever $Q \in \mathbb{F}_p[X]$ is taken irreducible:

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

(3) 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 $i^2 = -1$.

(4) 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:

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

To be more precise here, we assume of course that the characteristic of our model is $p = 2$, which reads $x + x = 0$ for any x, and so determines the addition table. As for the multiplication table, this is uniquely determined by $a^2 = -a - 1 = a + 1$.

(5) 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 $i^2 = -1$.

(6) Finally, at $q = p^k$ with $k \geq 3$ things become more complicated, but the main idea remains the same. We have for instance models for $\mathbb{F}_8$, $\mathbb{F}_{27}$ using $Q = X^3 - X - 1$, and a model for $\mathbb{F}_{16}$ using $Q = X^4 + X + 1$. Many other things can be said here. □

As another application of the above, which motivated Galois, we have:

THEOREM 4.22. *Unlike in degree $N \leq 4$, there is no formula for the roots of polynomials of degree $N = 5$ and higher, with the reason for this, coming from Galois theory, being that S_5 is not solvable. The simplest numeric example is $P = X^5 - X - 1$.*

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

(1) The first assertion, for generic polynomials, is due to Abel-Ruffini, but Galois theory helps in better understanding this, and comes with a number of bonus points too, namely the possibility of formulating a finer result, with Abel-Ruffini's original "generic", which was something algebraic, being now replaced by an analytic "generic", and also with the possibility of dealing with concrete polynomials, such as $P = X^5 - X - 1$.

(2) Regarding now the details of the Galois proof of the Abel-Ruffini theorem, assume that the roots of a polynomial $P \in F[X]$ can be computed by using iterated roots, a bit as for the degree 2 equation, or for the degree 3 and 4 equations, via Cardano. Then, algebrically speaking, this gives rise to a tower of fields as folows, with $F_0 = F$, and each F_{i+1} being obtained from F_i by adding a root, $F_{i+1} = F_i(x_i)$, with $x_i^{n_i} \in F_i$:

$$F_0 \subset F_1 \subset \ldots \subset F_k$$

(3) In order for Galois theory to apply well to this situation, we must make all the extensions normal, which amounts in replacing each $F_{i+1} = F_i(x_i)$ by its extension $K_i(x_i)$, with K_i extending F_i by adding a n_i-th root of unity. Thus, with this replacement, we can assume that the tower in (2) in normal, meaning that all Galois groups are cyclic.

(4) Now by Galois theory, at the level of the corresponding Galois groups we obtain a tower of groups as follows as follows, which is a resolution of the last group G_k, the Galois group of P, in the sense of group theory, in the sense that all quotients are cyclic:

$$G_1 \subset G_2 \subset \ldots \subset G_k$$

As a conclusion, Galois theory tells us that if the roots of a polynomial $P \in F[X]$ can be computed by using iterated roots, then its Galois group $G = G_k$ must be solvable.

(5) In the generic case, the conclusion is that Galois theory tells us that, in order for all polynomials of degree 5 to be solvable, via square roots, the group S_5, which appears there as Galois group, must be solvable, in the sense of group theory. But this is wrong, because the alternating subgroup $A_5 \subset S_5$ is simple, and therefore not solvable.

(6) Finally, regarding the polynomial $P = X^5 - X - 1$, some elementary computations here, based on arithmetic over $\mathbb{F}_2, \mathbb{F}_3$, and involving various cycles of length $2, 3, 5$, show that its Galois group is S_5. Thus, we have our counterexample.

(7) To be more precise, our polynomial factorizes over $\mathbb{F}_2$ as follows:

$$X^5 - X - 1 = (X^2 + X + 1)(X^3 + X^2 + 1)$$

We deduce from this the existence of an element $\tau\sigma \in G \subset S_5$, with $\tau \in S_5$ being a transposition, and with $\sigma \in S_5$ being a 3-cycle, disjoint from it. Thus, we have:

$$\tau = (\tau\sigma)^3 \in G$$

(8) On the other hand since $P = X^5 - X - 1$ is irreducible over $\mathbb{F}_5$, we have as well available a certain 5-cycle $\rho \in G$. Now since $< \tau, \rho >= S_5$, we conclude that the Galois group of P is full, $G = S_5$, and by (4) and (5) we have our counterexample.

(9) Finally, as mentioned in (1), all this shows as well that a random polynomial of degree 5 or higher is not solvable by square roots, and with this being an elementary consequence of the main result from (5), via some standard analysis arguments. □

4e. Exercises

Exercises:

Exercise 4.23.

Exercise 4.24.

Exercise 4.25.

Exercise 4.26.

Exercise 4.27.

Exercise 4.28.

Exercise 4.29.

Exercise 4.30.

Bonus exercise.

Part II

Surfaces, algebra

Kirkby, Kearsley, Keighley, Maghull
Harrogate, Huddersfield, Oldham, Lancs
Grimsby, Glossop, Hebden Bridge
It's Grim Up North

CHAPTER 5

Surfaces, manifolds

5a. Surfaces, quadrics

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

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

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

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

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 $C \subset \mathbb{R}^2$ in chapter 2, we have the following result:

THEOREM 5.3. *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. □

5b. Higher hypersurfaces

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

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

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

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

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 5.5 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 5.6. *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. □

5c. Manifolds, examples

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 $S_{\mathbb{R}}^{N-1} \subset \mathbb{R}^N$ 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 5.7. *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:

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

(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 5.8. *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 2.7. 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 5.7, we conclude that this formula holds in fact for any matrix A, by continuity, as desired. □

5d. Arbitrary fields

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

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

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 5.8, as follows:

THEOREM 5.10. *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 5.7 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 5.8 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:

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

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.

5e. Exercises

Exercises:

EXERCISE 5.11.

EXERCISE 5.12.

EXERCISE 5.13.

EXERCISE 5.14.

EXERCISE 5.15.

EXERCISE 5.16.

EXERCISE 5.17.

EXERCISE 5.18.

Bonus exercise.

CHAPTER 6

Abstract algebra

6a. Abstract algebra

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.

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

DEFINITION 6.1. *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 $a(b+c)=ab+ac$.*

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

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

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. □

The above result is quite interesting, obviously in relation with arithmetic, and suggests looking into the intermediate fields of numbers, as follows:

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

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 6.3. *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, when trying to draw the graph of f, 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 6.4. *$\mathbb{Q}$ 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 6.2, 6.3, 6.4 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 6.5. *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 6.6. *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 6.4, we have:

THEOREM 6.7. *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. □

6b. Rings and modules

Many other things can be said about fields, and we will be back to this in chapter 3, 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 6.8. *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 6.9. *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 6.10. *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 6.7, we have:

THEOREM 6.11. *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 6.10 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, later on in this book. 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 6.8, 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 6.12. *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 6.12. 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 6.10 too:

THEOREM 6.13. *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 6.14. *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:

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

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 chapter 5. There is an interesting link there with the notion of ideal, coming from:

PROPOSITION 6.15. *Given an arbitrary algebraic manifold, appearing as*

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

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

(1) *Any linear combination $P = \sum \lambda_i P_i$ vanishes on X.*
(2) *More generally, any combination $P = \sum P_i Q_i$ 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:

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

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 6.16. *The algebraic manifolds are precisely the sets of the form*

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

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

PROOF. In one sense, this comes from Proposition 6.15, 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 6.17 (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 $\lambda_i \in R$:

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

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

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

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 6.17 tells us that in our original definition of manifolds we can always assume that the family of polynomials $\{P_i\}$ there is finite. Equivalently, in our reformulation from Proposition 6.16, 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 6.17.

However, Theorem 6.17 is best remembered geometrically, as follows:

THEOREM 6.18. *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 something self-explanatory, coming from Theorem 6.17, and written somehow in the spirit of Proposition 6.16, with many other formulations being of course possible, and with meditating a bit on all this being a useful exercise for you. □

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

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

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

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

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. Obviously, what happens is that we have an inclusion $I \subset I_{X_I}$, which is not an isomorphism in general, due for instance to polynomials P satisfying $P \notin I$, but $P^n \in I$ for some $n \in \mathbb{N}$.

But this suggests doing some abstract algebra, as follows:

PROPOSITION 6.20. *Given an ideal $I \subset R$, define its radical as being:*

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

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 6.17, follows indeed from (3). □

We can now go back to the correspondences in Proposition 6.19, with the following key addition to the material there:

THEOREM 6.21. *Given an algebraic manifold $X \subset F^N$, its ideal, given by*

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

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 6.22. *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 algebraically 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 6.23 (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. $\square$

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 6.24 (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 $f_i \in I$:

$$1 = f_0 + f_1 x_{N+1} + \ldots + f_m x_{N+1}^m$$

(5) Now by multiplying by P^m, we obtain from this the following formula:

$$P^m = P^m f_0 + P^{m-1} f_1 + \ldots + f_m$$

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 6.25. *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, later in this book.

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 6.26. *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 6.27. *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 all this, later.

6e. Exercises

Exercises:

EXERCISE 6.28.

EXERCISE 6.29.

EXERCISE 6.30.

EXERCISE 6.31.

EXERCISE 6.32.

EXERCISE 6.33.

EXERCISE 6.34.

EXERCISE 6.35.

Bonus exercise.

CHAPTER 7

7a.

7b.

7c.

7d.

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

8a.

8b.

8c.

8d.

8e. Exercises

Exercises:

Exercise 8.1.

Exercise 8.2.

Exercise 8.3.

Exercise 8.4.

Exercise 8.5.

Exercise 8.6.

Exercise 8.7.

Exercise 8.8.

Bonus exercise.

Part III

Algebraic manifolds

Ooh ooh aah aah sexy eyes
I'm gonna take you to paradise
Hey hey my my can't you see
You were born to dance with me

CHAPTER 9

9a.

9b.

9c.

9d.

9e. Exercises

Exercises:

Exercise 9.1.

Exercise 9.2.

Exercise 9.3.

Exercise 9.4.

Exercise 9.5.

Exercise 9.6.

Exercise 9.7.

Exercise 9.8.

Bonus exercise.

CHAPTER 10

10a.

10b.

10c.

10d.

10e. Exercises

Exercises:

Exercise 10.1.

Exercise 10.2.

Exercise 10.3.

Exercise 10.4.

Exercise 10.5.

Exercise 10.6.

Exercise 10.7.

Exercise 10.8.

Bonus exercise.

CHAPTER 11

11a.

11b.

11c.

11d.

11e. Exercises

Exercises:

Exercise 11.1.

Exercise 11.2.

Exercise 11.3.

Exercise 11.4.

Exercise 11.5.

Exercise 11.6.

Exercise 11.7.

Exercise 11.8.

Bonus exercise.

CHAPTER 12

12a.

12b.

12c.

12d.

12e. Exercises

Exercises:

Exercise 12.1.

Exercise 12.2.

Exercise 12.3.

Exercise 12.4.

Exercise 12.5.

Exercise 12.6.

Exercise 12.7.

Exercise 12.8.

Bonus exercise.

Part IV

Advanced aspects

She never drinks the water
Makes you order French champagne
Once you've had a taste of her
You'll never be the same

CHAPTER 13

13a.

13b.

13c.

13d.

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

14a.

14b.

14c.

14d.

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

15a.

15b.

15c.

15d.

15e. Exercises

Exercises:

Exercise 15.1.

Exercise 15.2.

Exercise 15.3.

Exercise 15.4.

Exercise 15.5.

Exercise 15.6.

Exercise 15.7.

Exercise 15.8.

Bonus exercise.

CHAPTER 16

16a.

16b.

16c.

16d.

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