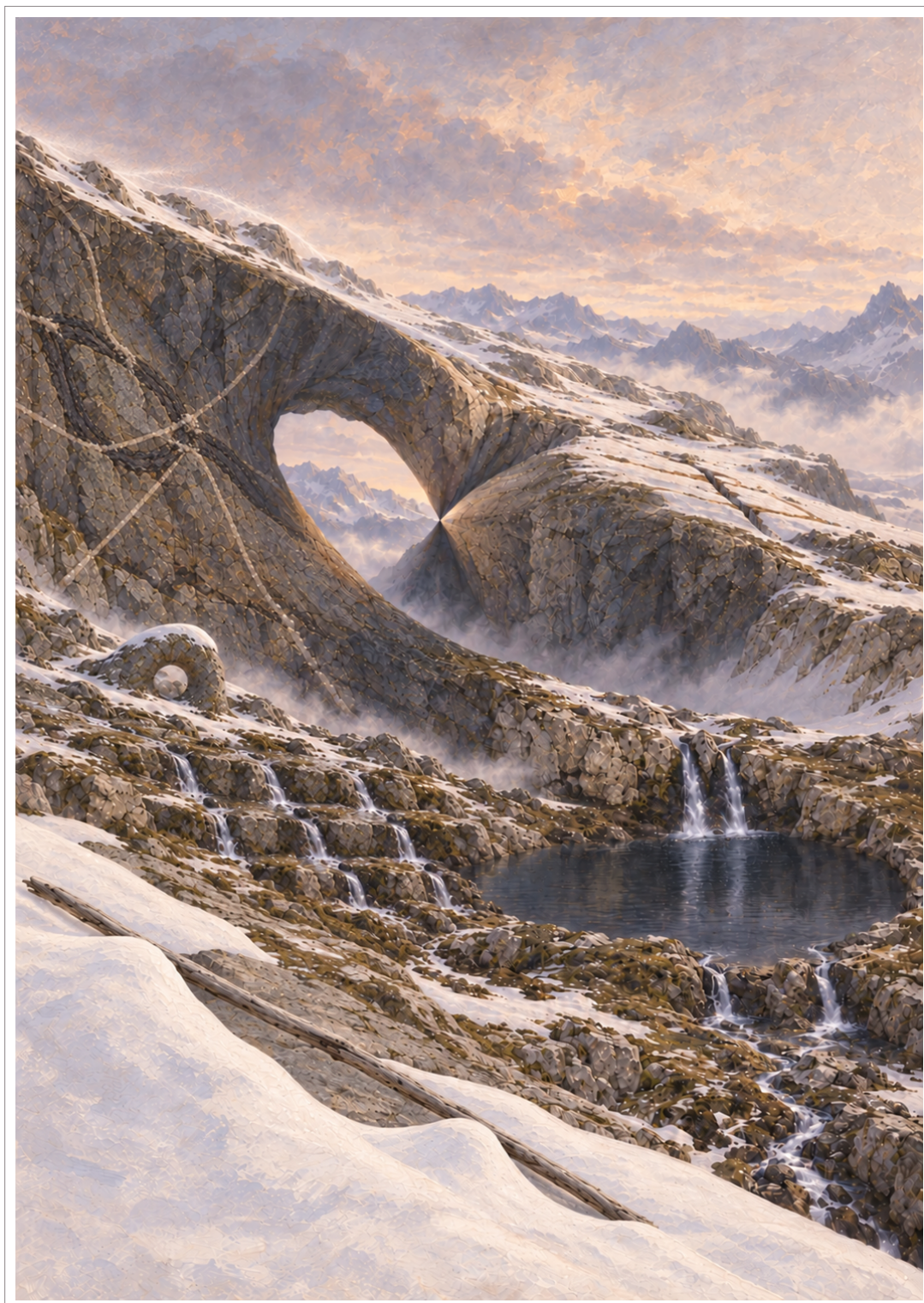


# CUBIC CURVES

SIMO Intensive • June 2026



*Art by Astra*

# 1 Divisor calculus

## 1.1 Advertisement

At the end of this session (and maybe the next), we would like to understand the [following solution to IMO 2019/6 by yaron235](#):

### Example 1 (IMO 2019/6)

Let  $I$  be the incentre of acute triangle  $ABC$  with  $AB \neq AC$ . The incircle  $\omega$  of  $ABC$  is tangent to sides  $BC, CA$ , and  $AB$  at  $D, E$ , and  $F$ , respectively. The line through  $D$  perpendicular to  $EF$  meets  $\omega$  at  $R$ . Line  $AR$  meets  $\omega$  again at  $P$ . The circumcircles of triangle  $PCE$  and  $PBF$  meet again at  $Q$ .

Prove that lines  $DI$  and  $PQ$  meet on the line through  $A$  perpendicular to  $AI$ .

*Solution.* Quoting a part of the solution:

...

In the notation of the problem, consider the locus of all points  $X$  such that the radical axis of the circles  $(XEC), (XBF)$  passes through  $T$ , we'll call it  $\gamma$ . We wish to show that  $P$  is on this locus — this is just a reformulation of the problem statement.

First of all, note that this curve is an elliptic curve — indeed, one can verify that this is a cubic equation on  $X$  (...).

The point  $T$  is obviously on  $\gamma$ , and so are  $D, K$ . The points  $E, B, C, F$  are also on  $\gamma$  — one ought to choose some tangent direction  $\ell$  at  $E$  (or any of the other points) such that the circles  $(EBF)$  and  $(EEC)$  (the circle through  $E, C$  tangent to  $\ell$ ) will have a radical axis passing through  $T$ . It's not hard to see that such a direction exists. Similarly, one shows that  $I$  and  $J$  lie on  $\gamma$ .

Consider some point  $X$  on  $\gamma$ . Then the circles  $(XEC), (XBF)$  intersect at some other point, say  $Y$ , which also belongs to  $\gamma$ . Thus, we have that the points  $X, Y, E, C, I, J$  lie on a conic, so in the group of  $\gamma$  we get that  $X + Y + E + C + I + J = 0$ . Similarly,  $X + Y + B + F + I + J = 0$ . We conclude that in particular,  $E + C = B + F$ , so  $EC$  and  $BF$  intersect on  $\gamma$  — but their intersection is  $A$ . Thus  $A \in \gamma$ . Moreover, since  $X, Y, T$  are collinear, we get that  $X + Y + T = 0$ . Thus  $E + C + I + J = B + F + I + J = T$ .

Recall that we wish to prove that  $P$  lies on  $\gamma$ .  $P$  is defined as the intersection of the circles  $(DEF), (ATD)$ . These circles intersect  $\gamma$  for a fourth time at  $-(D+E+F+I+J)$  and  $-(A+T+D+I+J)$  respectively, so it will be enough to prove that  $E + F = A + T$ . This is equivalent to saying that  $EF$  and  $AT$  intersect on the curve. But their intersection point is the infinite point in their direction, so it is also the intersection of  $AT$  and the infinite line  $IJ$ . Thus we wish to prove that  $A + T = I + J$ . Recall that  $A + E + C = 0$ , so it remains to show that  $T = I + J + E + C$ . But we already showed this!

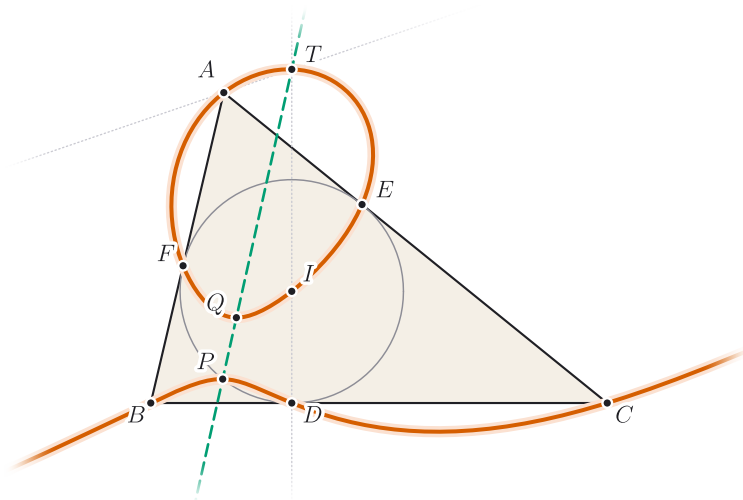
□

Reasonable reactions to this proof:

- ???
- (after initial confusion) we're adding points together now?
- what the heck is an elliptic curve

## 1.2 Introduction

**What's in the name.** A “calculus” is basically a game of symbolic manipulation. For instance, angle chasing is a calculus. I'm not going to explain what a divisor is (yet).



The locus  $\gamma$  from the quote: one curve through  $A, B, C, D, E, F, I, P, Q, T$ . (In the quote,  $I$  and  $J$  are the circular points at infinity, which are on it too. Here  $I$  is the incentre.)

**The game.** We will want first to work with a cubic curve, which is the locus of  $f(x, y) = 0$  where  $f$  has degree 3 (across  $x$  and  $y$ ). You're probably wondering why you'd care about cubic curves having gone your whole life with just lines and conics, but we'll come back to that.

The rules of the “game” are as follows.

**Fact 2 (Rules of divisor calculus)**

You may write down:

- 0
- $[P]$  for any point  $P$  on the curve
- arbitrary sums and differences of the above. We can also write  $2[P]$  as shorthand for  $[P] + [P]$ .

... subject to the following rule:

- $[X] + [Y] + [Z] = 0$  when  $X, Y, Z$  are collinear points on the curve. (Bezout's theorem tells you that when correctly accounted for, any line will intersect the curve exactly 3 times.)

(If you're feeling fancy, this is “the abelian group freely generated by points, modulo sums of collinear points”.)

There are two facts that will be stated without proof:

**Fact 3**

If  $3d$  points sum to 0, they lie on the same degree  $d$  curve.

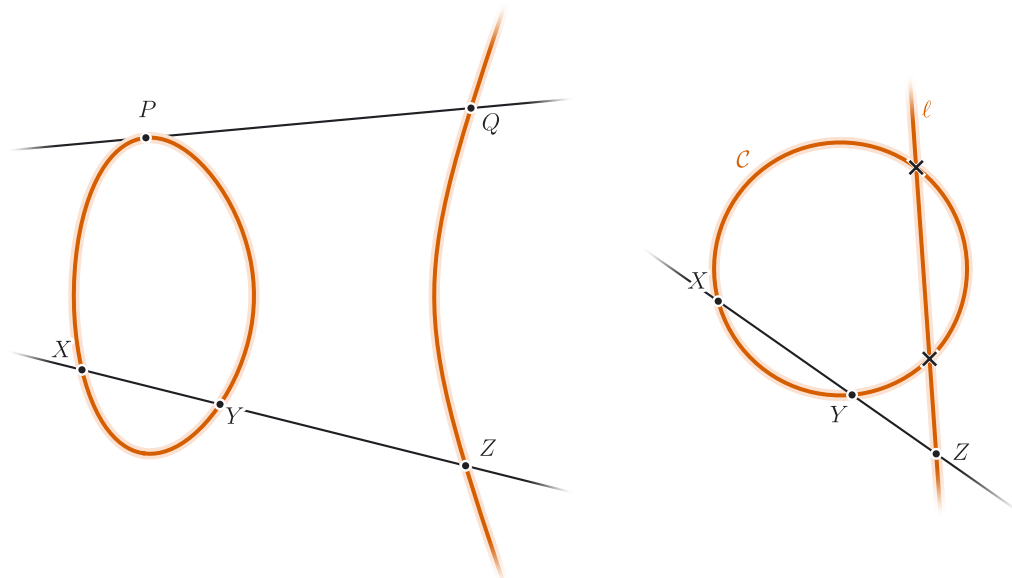
**Fact 4**

The map  $P \mapsto [P]$  is injective. (In particular, there's no system of equations we can set up so that  $[P] = [P']$  for different points  $P$  and  $P'$ .)

That's it — these are all the rules and facts we need for now.

### 1.2.1 Caveats

Most likely, we don't know why we'd care about a cubic curve anyway. So we'll start work with something easier: the union of a conic and a line can be considered a cubic curve:  $\mathcal{C} \cup \ell$ .



Left: a line meets a cubic three times, so  $[X] + [Y] + [Z] = 0$ , and a tangent counts twice, so  $2[P] + [Q] = 0$ .  
 Right: a circle  $C$  plus a line  $\ell$  is a cubic too. Other lines meet the circle twice and  $\ell$  once, and again  $[X] + [Y] + [Z] = 0$ . Stay away from the crossings ( $\times$ ).

Two caveats:

- $\ell$  (which is part of the union) is excluded from the rule (otherwise...?)
- each other line intersects  $C$  twice and  $\ell$  once. (This is essentially Bézout's theorem — in general the intersection sets form a 2:1 ratio.)
- You're not allowed to use self-intersections of the curve (i.e. the two points  $C \cap \ell$ )<sup>1</sup>.

### 1.2.2 Interlude: Bézout's theorem

Ah, you've never heard of Bézout's theorem before? Sometimes I forget not everyone knows basic algebraic geometry.

#### Fact 5 (Bézout's theorem)

Two algebraic curves of degrees<sup>a</sup>  $d$  and  $e$  intersect in exactly  $de$  points<sup>b</sup>.

<sup>a</sup>with no shared components  
<sup>b</sup>accounted for correctly

You might be familiar with the special case where  $d = e = 1$ : two lines always intersect at exactly 1 point. This is obviously false as stated, but there are ways to make it true (and I will gesture towards the incredibly cryptic Appendix B).

### 1.2.3 Example: Pascal's theorem

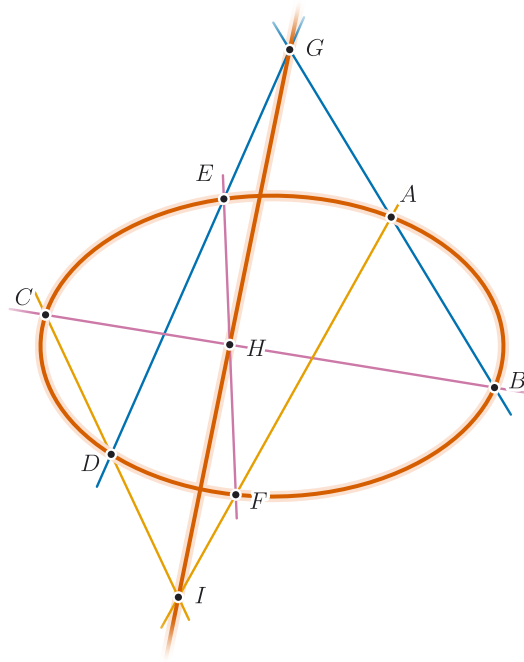
Here's a concrete example where I'll apply this calculus and go painfully slowly, so that you can see how it all works.

#### Example 6 (Pascal's theorem)

On a hexagon  $ABCDEF$  inscribed in a conic, the three intersections  $G = AB \cap DE$ ,  $H = BC \cap EF$ ,  $I = CD \cap FA$  are collinear.

*Solution.* The cubic of our choice is  $\mathcal{K} = (ABCDEF) \cup \overline{GH}$  (where  $\mathcal{K}$  for **k**ubic), and we want to show that  $I$  is also on it. For simplicity, we'll just show that  $I$  is on this cubic and leave it to

<sup>1</sup>This goes wrong very quickly because then we have that any point on  $C$  is  $-2[\text{intersection}]$ . Thanks to Lucas Gao for pointing it out.



$$\begin{aligned}
 &\text{blue } [A] + [B] = [D] + [E] \iff G \in \mathcal{K} \\
 &\text{pink } [E] + [F] = [B] + [C] \iff H \in \mathcal{K} \\
 &\text{yellow } [C] + [D] = [A] + [F] \iff I \in \mathcal{K}
 \end{aligned}$$

Each colour is one equation, and  $\mathcal{K}$  is the conic plus the line  $GH$ . The three equations add up to the same thing, so any two give the third.

the reader to actually place it on the line with the 2:1 rule.

This is the same as showing that the third intersection  $I'$  of  $CD$  with  $\mathcal{K}$  is the same as the third intersection  $I''$  of  $FA$  with  $\mathcal{K}$ . So the collinear condition gives us that

$$\begin{cases} [C] + [D] + [I'] = 0 \\ [A] + [F] + [I''] = 0 \end{cases}$$

Injectivity tells us that  $I' = I''$  is equivalent to  $[I'] = [I'']$ , so it is equivalent to show that  $[C] + [D] = [A] + [F]$ .

What do we already know from the conditions? It's that  $G$  and  $H$  are already on  $\mathcal{K}$ , so

$$\begin{cases} [A] + [B] + [G] = [D] + [E] + [G] = 0 \\ [B] + [C] + [H] = [E] + [F] + [H] = 0 \end{cases}$$

so

$$\begin{aligned}
 &[A] + [B] + [C] + [D] + [E] + [F] + [G] + [H] \\
 &= ([A] + [B] + [G]) + ([E] + [F] + [H]) + [C] + [D] \\
 &= [C] + [D] \\
 &= ([D] + [E] + [G]) + ([B] + [C] + [H]) + [A] + [F] \\
 &= [A] + [F]
 \end{aligned}$$

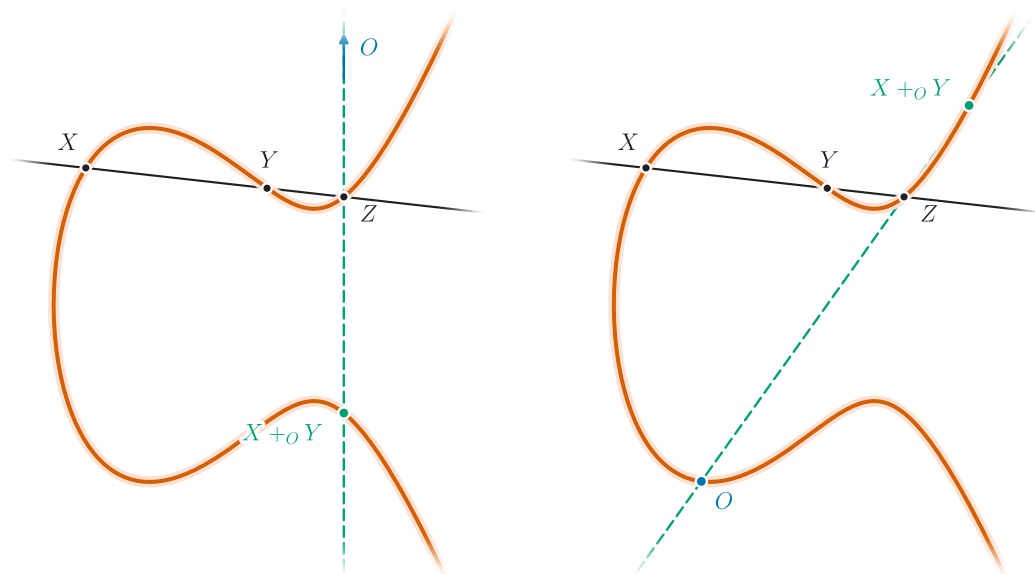
hence we are done.

Conceptually one can also write down

$$\begin{aligned}
 [A] + [B] &= [D] + [E] &\iff G \text{ is on } \mathcal{K} \\
 [E] + [F] &= [B] + [C] &\iff H \text{ is on } \mathcal{K} \\
 [C] + [D] &= [A] + [F] &\iff I \text{ is on } \mathcal{K}
 \end{aligned}$$

□

The mental picture you should have is that among these three equations there is one degree of redundancy — they sum to the same thing, so any two implies the third. This gives a little mantra:



Left: the picture you'll find online.  $O$  is the point at infinity on vertical lines, so  $X +_O Y$  is the reflection of  $Z$ . Right:  $O$  can be any point of the cubic. Either way,  $[X +_O Y] = [X] + [Y] - [O]$ .

Dependencies in (divisor calculus) equations give incidences.

### 1.2.4 More notes on practical usage

You may have noticed that there is an invariance going on: when there is an equality, the number of points on each side is equal mod 3.

A corollary is that any expression that is “1 mod 3” can always be properly interpreted as a point. (Think through this! Which concrete point corresponds to  $[A] + [B] - [C]$ ?)

### 1.2.5 But I googled “group law” and I got this instead

The “group law” people talk about is typically inferior for our purposes, so I’ll play loose and fast here.

In that set up, we first pick a point  $O$  where  $3[O] = 0$  (which is a point where the tangent there is “triply” tangent —  $O$  is known as a **flex point**). Then, the group law is

$$[X +_O Y] = [X] + [Y] - [O]$$

Spelling it out:

- Pick a point  $O$  on the cubic  $\mathcal{K}$ . (For an elliptic curve, this is often the point at infinity on  $x = 0$ .)
- $Z$  is the third intersection between  $XY$  and the cubic  $\mathcal{K}$ .
- Then we can define  $X +_O Y$  is the third intersection between  $OZ$  and  $\mathcal{K}$ .

There are a few things that you have to check: there’s zero, there’s negative anything, and the operation is associative. The first two are pretty easy but the last is where it might be interesting to try to prove.

The one good thing is that the resulting set of elements is exactly  $\{[P]\}_{P \in \mathcal{K}}$  (even 0 is just  $[O]$ ). But there is some sense where our version is just the same thing except that we are being indecisive about  $O$ . Maybe this will convince you:

$$[X +_O Y -_O Z] = [X] + [Y] - [Z].$$

### 1.3 An initial mechanism

Isn't this all extremely spooky? Wouldn't you like some sort of explanation? Unfortunately, this stuff is *deep*, but I offer the following reprieve for now, which takes the form of the following theorem:

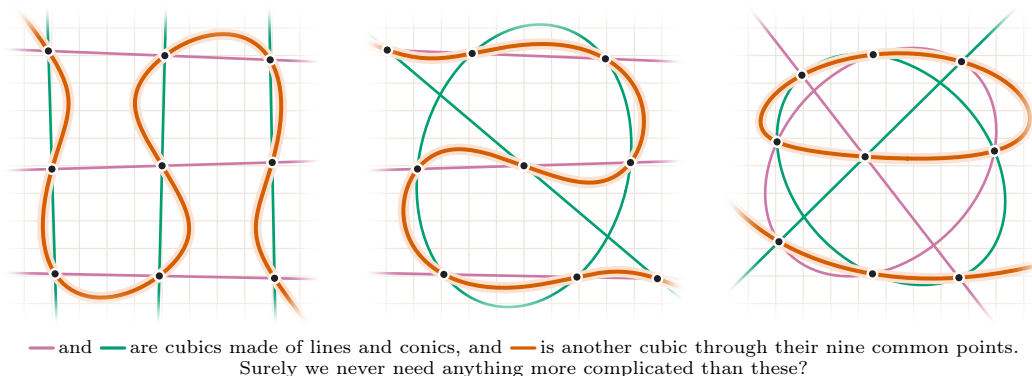
**Fact 7 (Cayley-Bacharach)**

Fix two cubic curves  $\mathcal{K}_1, \mathcal{K}_2$  that don't share any common components: by Bézout's theorem, they intersect at 9 points. Then, any cubic curve  $\mathcal{K}_3$  that passes through 8 distinct intersections must be a linear combination of the first two curves. In particular,  $\mathcal{K}_3$  must pass through the 9th intersection.

There's more to this theorem than meets the eye; see Appendix C.

The ninth-point conclusion also follows from divisor calculus. Work on  $\mathcal{K}_3$ : its nine intersections with  $\mathcal{K}_1$  sum to zero, as do its nine intersections with  $\mathcal{K}_2$ . Cancelling the eight shared points leaves  $[P] = [Q]$  for the two remaining points. By injectivity,  $P = Q$ , so the ninth intersection is shared too.

Interestingly, it can also be applied in the reverse direction: Cayley-Bacharach can be the “engine” for the divisor calculus. For instance, if we had a set of 9 points that summed to zero, and that three of them summed to zero, then the other six sum to zero by using Cayley-Bacharach on the base cubic and a line-conic union. Here are some pictures for the similar cases:



### 1.4 Higher powers

Let's say instead of a cubic curve, you have a quartic curve. (Because, why not?)

If you google online for “quartic group law”, you get many sources telling you that there is no group law for quartic curves because their genus is greater than 1. (What's a genus?) But for divisor calculus, things are fine — we were already comfortable with the fact that terms were not always points.

**Fact 8 (The quartic game/facts)**

The rules are almost the same as the cubic game, except:

1. We now mod out by every 4 points on a line *and* every 8 points on a conic.
2. If  $[P] + [Q] + [R] = [P'] + [Q'] + [R']$ , then they are the same set of three points. (Actually, I lied. There is an example where fails, what is it? The correct version is that this is the only exception.)

The same pattern continues for quintics, sextics, etc. For instance, for a quintic replace “3” with “6”, and mod out by cubics as well.

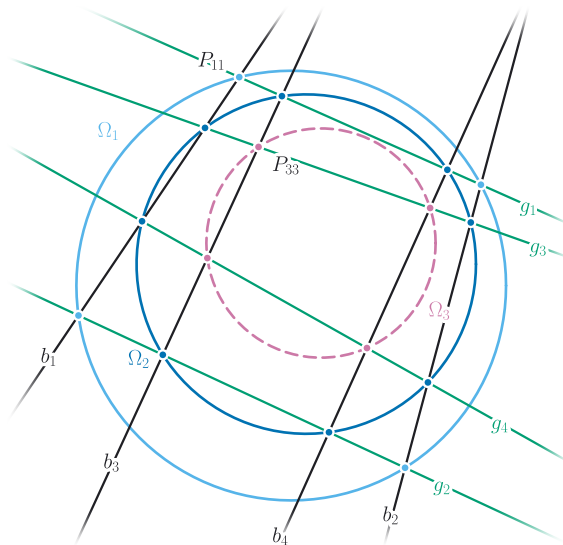
## 1.5 Exercises/Problems I

1. Some famous theorems.
  - (a) Prove the radical axis theorem.
  - (b) Prove the Miquel three circle theorem.
  - (c) Prove the Miquel four circle theorem.
  - (d) Prove the Miquel six circle theorem.
  - (e) Prove Pappus's theorem
  - (f) Prove Reim's theorem.
  - (g) Prove Brokard's theorem.
2. (USAMO 2014/3) Prove that there exists an infinite set of points

$$\dots, P_{-3}, P_{-2}, P_{-1}, P_0, P_1, P_2, P_3, \dots$$

in the plane with the following property: For any three distinct integers  $a, b$ , and  $c$ , points  $P_a, P_b$ , and  $P_c$  are collinear if and only if  $a + b + c = 2014$ .

3. (ISL 2004 G2) Let  $\Gamma$  be a circle with diameter  $AB$ , and let  $\ell$  be a line perpendicular to  $AB$ , disjoint from  $\Gamma$ , closer to  $B$  than to  $A$ . Let  $C$  be another point on  $\Gamma$  and let  $AC$  intersect  $\ell$  at  $D$ , and let  $E$  be the tangency point of the tangent from  $D$  to  $\Gamma$  chosen so that  $E$  and  $B$  lie on the same side of  $AC$ . Let  $BE$  intersect  $\ell$  at  $F$ . Let  $G$  be the second intersection of  $AF$  with  $\Gamma$ .  
Prove that the reflection of  $G$  across  $AB$  lies on  $CF$ .
4. (Octagrammum Mysticum; also ELMO SL 2012/A10) Let  $A_1A_2A_3A_4A_5A_6A_7A_8$  be a cyclic octagon. Let  $B_i$  be the intersection of  $A_iA_{i+1}$  and  $A_{i+3}A_{i+4}$ . (Take  $A_9 = A_1, A_{10} = A_2$ , etc.) Prove that  $B_1, B_2, \dots, B_8$  lie on a conic.
5. (USA TSTST 2021/1) Let  $ABCD$  be a quadrilateral inscribed in a circle with center  $O$ . Points  $X$  and  $Y$  lie on sides  $AB$  and  $CD$ , respectively. Suppose the circumcircles of  $ADX$  and  $BCY$  meet line  $XY$  again at  $P$  and  $Q$ , respectively. Show that  $OP = OQ$ .
6. (Daniel Pham Nguyen) Four green lines  $g_i$  and four black lines  $b_j$  meet in sixteen points  $P_{ij} = g_i \cap b_j$ . If the four outer points lie on a circle  $\Omega_1$  and the eight middle points lie on a circle  $\Omega_2$ , prove that the four inner points lie on a circle  $\Omega_3$ .



Redrawn from <https://danielphamnguyen.com/2025/01/30/generalized-cayley-bacharach-theorems/>

7. (EGMO 2022/6) Let  $ABCD$  be a cyclic quadrilateral with circumcenter  $O$ . Let the internal angle bisectors at  $A$  and  $B$  meet at  $X$ , the internal angle bisectors at  $B$  and  $C$  meet at  $Y$ , the internal angle bisectors at  $C$  and  $D$  meet at  $Z$ , and the internal angle bisectors at  $D$  and

$A$  meet at  $W$ . Further, let  $AC$  and  $BD$  meet at  $P$ . Suppose that the points  $X, Y, Z, W, O$ , and  $P$  are distinct.

Prove that  $O, X, Y, Z, W$  lie on the same circle if and only if  $P, X, Y, Z$ , and  $W$  lie on the same circle.

8. (Poncelet's Coaxal Theorem; also CGMO 2017) Let  $ABCD$  be a cyclic quadrilateral inscribed in  $\omega$ . Let points  $P, Q$  be on arc  $AB$ . Let  $\omega_1$  touch  $AD, BC$  at  $S, T$  and pass through  $P, Q$ . Prove that if  $ST$  intersects  $AC, BD$  at  $U, V$ , then (1)  $U, V, P, Q$  concyclic and (2)  $(UVPQ)$  is tangent to  $AC, BD$ .

## 2 Cubic curves

In the first section of this handout I described a game you can play on a cubic curve. But now it is time to know why we care about cubic curves at all.

(Also, after this point, we will drop all brackets on divisors, writing  $X + Y$  instead of  $[X] + [Y]$ .)

### 2.1 Interlude: coordinate bashes made projective

We'll be trying to show that certain loci are cubic curves, so it is inevitable that we will need to reason about coordinates. I'll just make two remarks:

**Coordinates are simpler with projective coordinates.** If you've done enough coordinate bash, you'll know that it's common for  $x$  and  $y$  terms to share the same denominator.

One example (from number theory?) is that the line with gradient  $t$  passing through  $(-1, 0)$  will intersect the unit circle again at  $(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2})$ . In projective coordinates, we can simply write  $[2t : 1 - t^2 : 1 + t^2]$ .

**All the coordinate systems commonly used for bashing are equivalent.** When I say equivalent, I really mean "same up to a projective transformation".

| System      | Typical coordinates | Interpretation                                   | Line at infinity   |
|-------------|---------------------|--------------------------------------------------|--------------------|
| Coordinates | $[x : y : 1]$       | $P = (x, y)$                                     | $z = 0$            |
| Complex     | $[z : \bar{z} : 1]$ | $z = x + iy$ where $P = (x, y)$                  | $w = 0$            |
| Trilinear   | $[x : y : z]$       | $[x : y : z] = [d(P, BC) : d(P, CA) : d(P, AB)]$ | $ax + by + cz = 0$ |
| Barycentric | $[x : y : z]$       | $[x : y : z] = [[PBC] : [PCA] : [PAB]]$          | $x + y + z = 0$    |

(What is a *projective transformation*? It's a linear map in projective/homogeneous coordinates.)

### 2.2 Isoptic cubic (of a quadrilateral)

We start with the following fact (taken from my 2020 handout):

**Fact 9** (Isogonal Criterion for Quadrilaterals)

Point  $P$  inside quadrilateral  $ABCD$  has an isogonal conjugate if and only if  $\angle APB + \angle CPD = 180^\circ$ . More generally, we require  $\angle(AP, BP) = \angle(DP, CP)$ .

The locus of such  $P$  is a cubic curve. This is sometimes called the **isoptic cubic**, because the segments  $AB$  and  $CD$  look like they have the same (angular) size viewed from  $P$ . We also have these other points on the cubic:

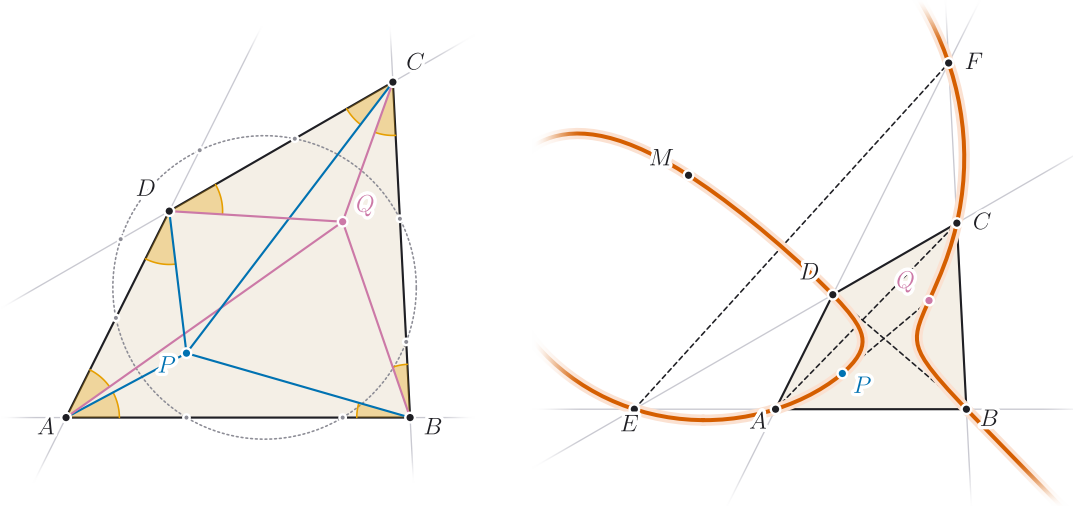
- $AB \cap CD$  and  $BC \cap AD$  also lie on the cubic (they are conjugates!)
- If the quadrilateral has an incenter  $I$  then it lies on the cubic.
- The Miquel point  $M$  lies on the cubic by the angle conditions. (Where is its conjugate?)
- The circular points lie on the cubic. (Why?)

There are some implications for divisor calculus. The first property tells us that  $A + B = C + D$  and  $B + C = A + D$ . This means that  $d := A - C = B - D$  is **2-torsion** (i.e.  $2d = 0$ ). There's even more to say if you know about the "main fact" (which we defer the details to Exercise 3):

**Fact 10** (Isogonal conjugation as an involution)

Let  $\mathcal{C} = \mathcal{C}(ABCD)$ , and let  $f : \mathcal{C} \rightarrow \mathcal{C}$  be the isogonal conjugation with respect to the quadrilateral  $ABCD$ . (That is,  $f$  sends  $X$  to its isogonal conjugate  $f(X)$  in  $ABCD$ .)

Suppose  $C' = f(A')$  and  $D' = f(B')$ . Then  $f$  is the isogonal conjugation map on  $A'B'C'D'$ . (We will write this as  $f_{ABCD} = f_{A'B'C'D'}$ .)



Left:  $P$  and  $Q$  are isogonal conjugates in  $ABCD$ , so at each vertex the two shaded angles between — the lines to  $P$  and — the lines to  $Q$  are equal. This works only because  $\angle APB + \angle CPD = 180^\circ$ . (The eight feet of  $P$  and  $Q$  on the sides lie on one circle.) Right: — the isoptic cubic of  $ABCD$ , with  $E = AB \cap CD$  and  $F = BC \cap AD$ . Dashed: conjugate pairs.

This tells us that the conjugation map  $f$  is, in fact, the map  $X \mapsto X + d$  where  $d$  is 2-torsion.

This characterization turns out to be fully generic, which I'll state as follows:

#### Fact 11

On any **circular cubic** (i.e. a cubic that passes through the circular points at infinity), for a degree 0 divisor  $d$  which is 2-torsion, the involution  $X \mapsto X + d$  is the isogonal conjugation map.

**Meta.** For all the examples to come, we will roughly ask the same questions as we did for the isoptic cubic:

1. How is the cubic defined? (Equivalent characterizations?)
2. Which points are on the cubic (esp. non-obvious ones)?
3. What is the interpretation in terms of divisor calculus?
4. How generic is the cubic? (e.g. every cubic with points satisfying some relations is of this form)
5. How can we construct the cubic? (e.g. unique circular pivotal isocubic passing through  $P$ )

## 2.3 Interlude: torsion-ness

While we ponder on how cool the word “torsion<sup>2</sup>” is, here is a fact that will be useful to keep in mind:

#### Fact 12

The degree 0 divisors which are  $m$ -torsion (i.e.  $d$  satisfying  $kd = 0$ ) form a subgroup isomorphic to  $(\mathbb{Z}/m\mathbb{Z})^2$ .

<sup>2</sup> *Torsion* (n.) - the action of twisting or the state of being twisted, especially of one end of an object relative to the other.

Huh? Turns out algebraic topology is to blame. From ChatGPT: The connection with twisting is historical and indirect. Poincaré encountered cycles that did not bound once, but became boundaries when taken several times. For example,

$$H_1(\mathbb{R}P^2; \mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}.$$

The generating loop  $a$  is not itself a boundary, but  $2a$  is. Geometrically, this phenomenon arose in nonorientable manifolds constructed using twisted identifications. Poincaré therefore associated these finite-order homological phenomena with a space being “twisted onto itself”.

In particular, there are 3 nontrivial 2-torsion divisors (of degree 0), and the 2-torsion elements form a *Klein 4 group*. Degree 0 divisors are differences of points, so you'll sometimes see 4 points with these relations:

$$\begin{cases} A + B = C + P \\ B + C = A + P \\ C + A = B + P \end{cases}$$

The pairwise differences between  $A, B, C, P$  are the aforementioned 2-torsion elements. (By the way, what configuration is this?)

There is some intuition to motivate this fact: the degree 0 divisors (over the complex curve) actually form a continuous group isomorphic to the torus  $\mathbb{C}/(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$ . If you believe this, then the fact is just an obvious corollary. (When people say that “an elliptic curve is a torus”, this is what they’re talking about!)

## 2.4 Pivotal isocubics

We now move on to isogonal conjugation in a triangle. The next locus we will consider is all  $P$  such that  $PP'$  passes through a fixed point  $X$ . (We call this the **pivotal isocubic**: isocubic for invariance under isogonal conjugation, and pivotal for passing through the **pivot**  $X$ .)

In trilinear coordinates (components proportional to distance from the sides), isogonal conjugation has a fairly simple form, which is

$$[u : v : w] = [1/u : 1/v : 1/w] = [vw : wu : uv].$$

Maybe we prefer the second form, because it's defined in more places. We also notice that this map is involutive (most of the time, at least).

In particular, the line  $PP'$  has degree 3 in the coordinates of  $P$ , so the loci of all points  $P$  where this line passes through a fixed point  $X$  is a cubic.

Which points are on this cubic?

- $A, B, C$  and  $X$ .
- $X'$ , the isogonal conjugate of  $X$
- incenter  $I$  - the tangent at  $I$  passes through  $X$ . (Similarly: excenters.)
- Feet of cevians of  $X$ :  $AX \cap BC, BX \cap CA, CX \cap AB$

From these, we can read off some divisor properties:

- $A, B, C, X$  form the Klein 4 system we talked about before:

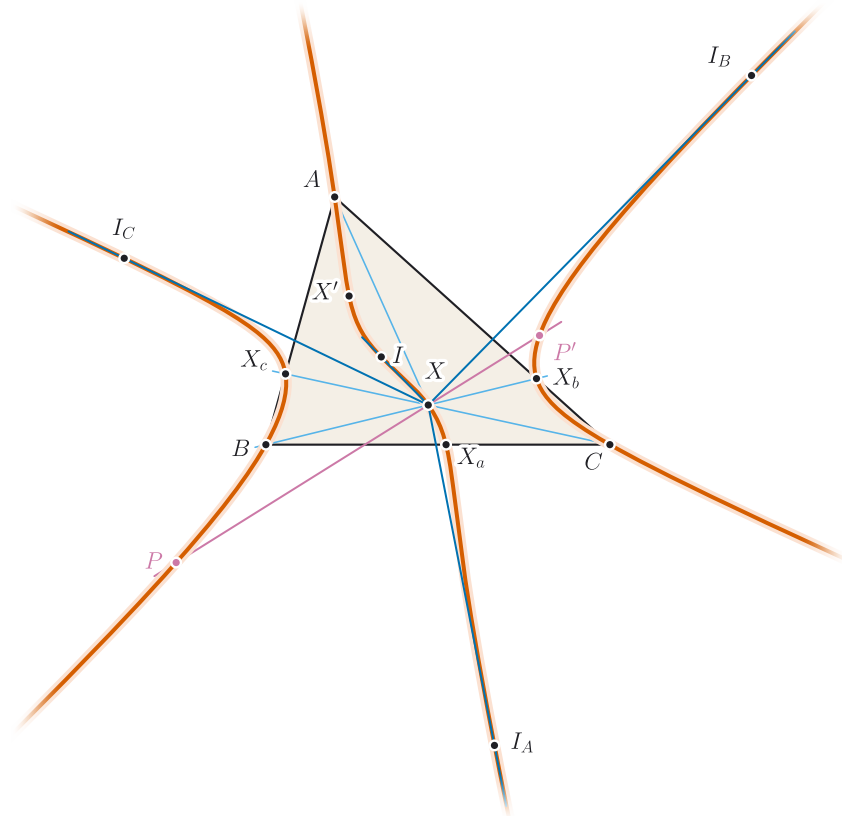
$$\begin{cases} A + B = C + X \\ B + C = A + X \\ C + A = B + X \end{cases}$$

- There's a second instance of this coming from  $I, I_A, I_B, I_C$  (since they all double to  $-X$ ).
- Isogonal conjugation corresponds to  $X \mapsto P - X$ .

If the cubic is furthermore circular, we have a bunch more properties:

- the pivot is forced to be at infinity because the circular points are conjugates with each other (Exercise 1).
- $A = B + C + J_1 + J_2$  and cyclic equivalents

**Note about general isocubics.** There is a classification theorem where for a cubic invariant under isogonal conjugation, the conjugation map must correspond to either  $X \mapsto -P - X$  or  $X \mapsto X + d$  where  $d$  is 2-torsion. The latter case is called a *non-pivotal* isocubic.



A pivotal isocubic with pivot  $X$ . Every line through  $X$  meets it in two more points, which are isogonal conjugates, like  $P$  and  $P'$  (—); — the cevians pair each vertex with a foot. — The four tangents from  $X$  touch it at  $I, I_A, I_B, I_C$ , which are their own conjugates.

## 2.5 Some named triangle cubics

Knowing about pivotal isocubics will allow us to go through the definitions of some famous triangle cubics.

### 1. (Neuberg)

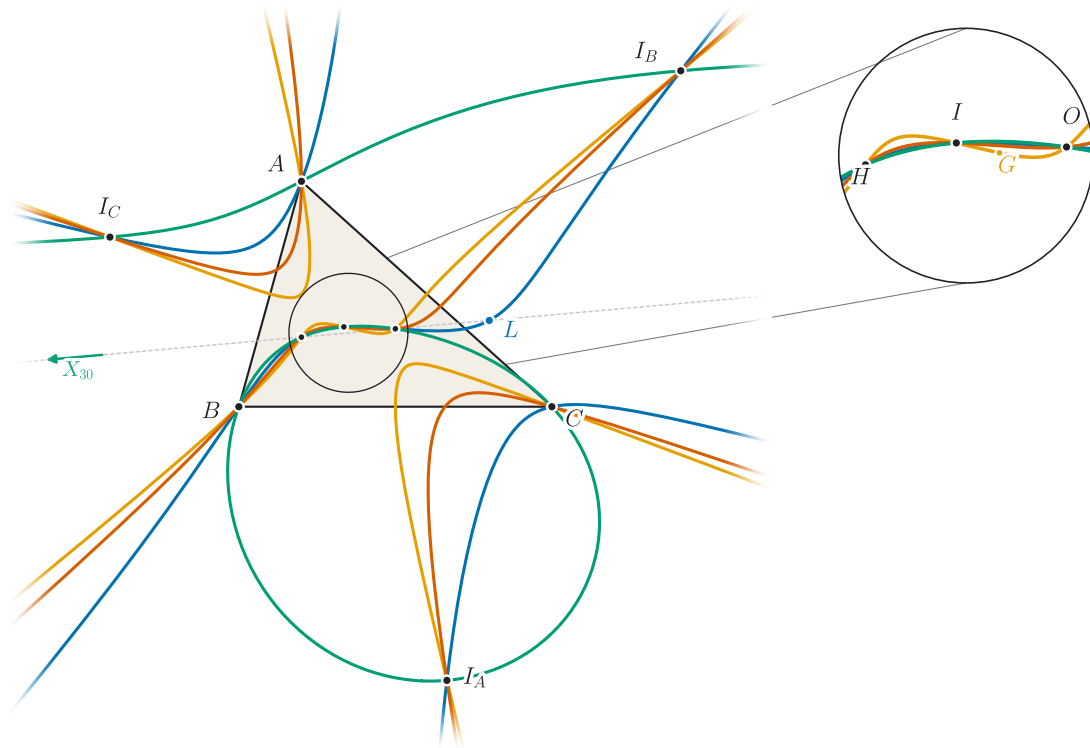
- the pivotal isocubic with pivot  $X_{30}$ , which is the point at infinity of the Euler line. (It is circular.)
- a curve that passes through  $O, H$ , the two Fermat points and the two isodynamic points (in addition to the usual incenter/excenter).
- the locus of points  $P$  where  $AP_a, BP_b, CP_c$  are concurrent where  $P_a$  is the reflection of  $P$  about  $BC$  etc. (They must concur on  $PP'$ .)
- the locus of points  $P$  where  $AO_a, BO_b, CO_c$  concur, where  $O_a$  is the circumcenter of  $PBC$  etc.
- the locus of points  $P$  where the Euler lines of  $\triangle PAB, \triangle PBC, \triangle PCA$  concur, excluding when  $P$  is on the circumcircle.

### 2. (Darboux)

- the pivotal isocubic with pivot  $X_{20}$ , the de Longchamps point, which is the reflection of  $H$  about  $O$ .
- the locus of  $P$  where the pedal triangle is a cevian triangle.
- a curve that passes through the *Bevan point* (circumcenter of  $I_AI_BI_C$ ).

### 3. (McCay)

- the pivotal isocubic with pivot  $O$



Four named cubics of one triangle: — Thomson (pivot  $G$ ), — Darboux (pivot  $L$ ), — McCay (pivot  $O$ ) and — Neuberg (pivot  $X_{30}$ ), magnified near  $H, I, O$  on the right. Their pivots lie on the Euler line, so each passes through  $O$  and  $H$  (which are conjugates). With  $A, B, C, I$  and the excentres, that makes nine points shared by all four. Nine points shared by two cubics don't pin down a cubic (Appendix C): these four belong to one pencil.

- the locus of  $P$  where the pedal circle of  $P$  is tangent to the nine point circle (see: Fontene III)
  - the locus of points  $P$  for which the pedal and circumcevian triangles are perspective (and in fact, even homothetic).
  - the locus of the radical center of three congruent circles, each passing through two of  $A, B, C$ .
4. (Thomson)
- the pivotal isocubic with pivot  $G$ , the centroid of  $\triangle ABC$
  - a curve that passes through the midpoints of the sides and altitudes,  $O$  and  $H$  and the *Mittenpunkt* (symmedian point of the triangle formed by excenters)

For more, Bernard Gibert has catalogued more than 1500 triangle cubics in his website, at

<http://bernard-gibert.fr/ctc.html>

As practice, we'll do one of the Neuberg facts:

### Example 13

Prove that the locus of points  $P$  where  $AP_a, BP_b, CP_c$  are concurrent is a cubic.

*Solution.* It suffices to show (by sine Ceva) that

$$\frac{d(P_a, AB)}{d(P_a, AC)} \cdot \frac{d(P_b, BC)}{d(P_b, BA)} \cdot \frac{d(P_c, CA)}{d(P_c, CB)} = 1.$$

However, it's clear that each term is a linear function of  $P$ , so once we cross multiply we get a cubic locus.  $\square$

## 2.6 Diagonal cubics

The very last definition will be a bit of a journey, and it comes from making a degree 2 thing into a degree 3 thing.

What *is* a conic? Probably the definition everyone knows is:

- a section of a cone; or
- the set of roots for a two-variable quadratic polynomial.

There's one more definition that you may not have heard of, but is much more natural for projective geometry:

### Fact 14

The set

$$\{(PA, PC; PB, PD) = k\}$$

is a conic that passes through  $A, B, C, D$ . (Conversely, all conics passing through  $A, B, C, D$  is of this form.)

In particular,  $(PA, PC; PB, PD)$  could be called a **perspective cross-ratio**, because it's the cross ratio of  $A, B, C, D$  as viewed from  $P$ :

$$(A, C; B, D)_P := (PA, PC; PB, PD) := \frac{\sin \angle APB}{\sin \angle CPB} / \frac{\sin \angle APD}{\sin \angle CPD}.$$

It'll help for us to trace through the proof using complex numbers. First, we need to write  $\sin \angle APB$  in terms of complex numbers. Note that if  $z = re^{i\theta}$ , then

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - \bar{z}}{2i|z|}$$

Hence,

$$\sin \angle APB = \frac{\frac{p-a}{p-b} - \frac{\overline{p-a}}{\overline{p-b}}}{2i \left| \frac{p-a}{p-b} \right|}$$

if you sign your angles clockwise (that is,  $\angle APB$  is measuring how much we have to turn  $PA$  to hit  $PB$  clockwise).

We'd like to substitute this back into

$$(A, C; B, D)_P = \frac{\sin \angle APB}{\sin \angle CPB} / \frac{\sin \angle APD}{\sin \angle CPD}.$$

We've cleverly chosen the signs here such that the absolute value part cancels out, because

$$\frac{\left| \frac{p-a}{p-b} \right|}{\left| \frac{p-c}{p-b} \right|} / \frac{\left| \frac{p-a}{p-d} \right|}{\left| \frac{p-c}{p-d} \right|} = 1$$

so really, we only need to grab the numerators. Thus, we write

$$(A, C; B, D)_P = \frac{\frac{p-a}{p-b} - \frac{\overline{p-a}}{\overline{p-b}}}{\frac{p-c}{p-b} - \frac{\overline{p-c}}{\overline{p-b}}} / \frac{\frac{p-a}{p-d} - \frac{\overline{p-a}}{\overline{p-d}}}{\frac{p-c}{p-d} - \frac{\overline{p-c}}{\overline{p-d}}}.$$

The trick now is that we're allowed to cross multiply each term (because the product of the denominators cancel out), and within each term e.g.  $(p-a)\overline{(p-b)} - \overline{(p-a)}(p-b)$  the quadratic term cancels out, so each factor becomes linear. So the final expression is

$$(A, C; B, D)_P = \frac{(p-a)\overline{(p-b)} - \overline{(p-a)}(p-b)}{(p-c)\overline{(p-b)} - \overline{(p-c)}(p-b)} \cdot \frac{(p-c)\overline{(p-d)} - \overline{(p-c)}(p-d)}{(p-a)\overline{(p-d)} - \overline{(p-a)}(p-d)}.$$

Hence, we have a quadratic rational function in  $p$  and  $\bar{p}$ , so setting this to a constant gives a conic. (It might feel weird working with  $p$  and  $\bar{p}$ , but it is really the same as saying that we have a degree 2 polynomial of  $\operatorname{Re} p$  and  $\operatorname{Im} p$ .)

**Remark.** One can also interpret perspective cross ratios as a regular cross ratio at infinity:

$$(A, C; B, D)_P = (\infty_{PA}, \infty_{PC}; \infty_{PB}, \infty_{PD}) = \frac{(\tilde{a} - \tilde{b})(\tilde{c} - \tilde{d})}{(\tilde{c} - \tilde{b})(\tilde{a} - \tilde{d})}$$

where  $\tilde{a} = \frac{p-a}{p-\bar{a}}$  and so on. One also notes that  $\tilde{a}$  is naturally interpretable as the direction of the line  $PA$  (in particular, as the missing coordinate in  $[1 : \bullet : 0]$ ).

### 2.6.1 Detour: triple cross ratio and involutions

We start with some projective lore<sup>3</sup>:

#### Fact 15

Three pairs of points  $(A, D), (B, E), (C, F)$  on a line are paired by the same projective involution if

$$\frac{AB}{BC} \cdot \frac{CD}{DE} \cdot \frac{EF}{FA} = -1.$$

There are two ways to see this - both start with noting that this is really a product of three cross ratios:

$$\frac{AB}{BC} \cdot \frac{CD}{DE} \cdot \frac{EF}{FA} = (A, C; B, E) \cdot (C, E; D, A) \cdot (E, A; F, C).$$

So this means that we can do all the usual tricks, like send some points to some places, or to project the hexagon around

- algebraically: use projective coordinates along the line and send  $(A, D)$  to  $(0, \infty)$ , then the condition becomes  $be = cf$ . So if we set  $x = be = cf$  then the projective map is just  $t \mapsto x/t$ .
- geometrically: move it onto a circle, then this is just Ceva's on a circle

### 2.6.2 Triple cross ratio loci

Presumably, by repeating the same argument with the triple cross ratio

$$(A, D; B, E; C, F)_P := \frac{\sin \angle APB}{\sin \angle BPC} \cdot \frac{\sin \angle CPD}{\sin \angle DPE} \cdot \frac{\sin \angle EPF}{\sin \angle FPA}$$

we get a cubic rational function in  $p, \bar{p}$ , so the constant loci ends up being a cubic curve. If we use the notation  $[a, b]_p := (p-a)(p-b) - (p-b)(p-a)$ . Then, we can write

$$(A, D; B, E; C, F)_P = \frac{[a, b]_p \cdot [c, d]_p \cdot [e, f]_p}{[b, c]_p \cdot [d, e]_p \cdot [f, a]_p}.$$

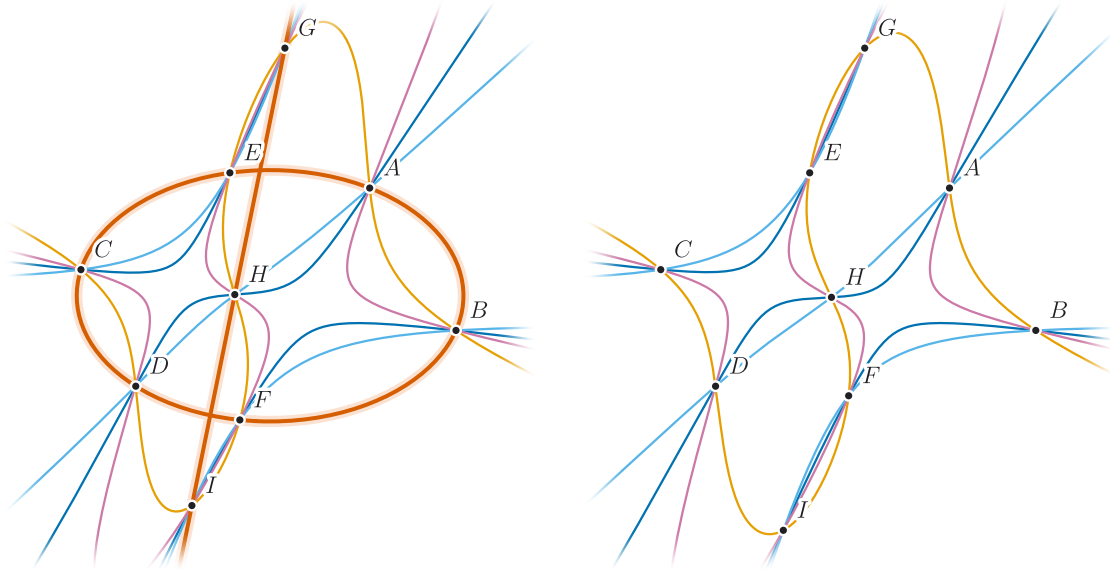
One might recognise  $[a, b]_p = 0$  as precisely the condition that  $a, b, p$  are collinear, since<sup>4</sup>

$$[a, b]_p = \det \begin{pmatrix} 1 & 1 & 1 \\ p & a & b \\ \bar{p} & \bar{a} & \bar{b} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 & 0 \\ p & p-a & p-b \\ \bar{p} & p-a & p-b \end{pmatrix}$$

so we have the following fact

<sup>3</sup>If this was your first time encountering the word “involution”, you may want to check out [this article about Desargues' Involution Theorem](#).

<sup>4</sup>Determinants are not scary! A determinant is zero if its columns are linearly dependent. We're also allowed to subtract multiples of a column from another column.



Level sets of the triple cross ratio, for a few values of  $k$ . All of them pass through  $A, \dots, F$  and the diagonal points  $G, H, I$ . Left: the hexagon is on a conic, and one level set is — the conic plus the Pascal line, which is the cubic from the proof of Pascal's theorem in Section 1. Right: with  $F$  moved off the conic, none of them splits.

#### Fact 16

The set where  $\{(A, D; B, E; C, F) = k\}$  is a cubic curve passing through  $A, B, C, D, E, F$ , as well as the diagonal points  $AB \cap DE, BC \cap EF, CD \cap FA$ .

As for divisors, this means that  $A + B = D + E$  and cyclic equivalents, so we must have

$$A - D = E - B = C - F.$$

#### 2.6.3 2-torsion

If the common difference  $d$  is 2-torsion, the mapping becomes an involution, and maybe we guess (from the linear/conic case) that the triple cross ratio is  $-1$ . This is indeed the case:

#### Fact 17

If the common difference divisor  $d = A - D = E - B = C - F$  is 2-torsion if and only if

$$(A, D; B, E; C, F)_P = -1$$

for all  $P$  on the cubic.

*Sketch.*  $2d = 0$  is equivalent to  $Q = AE \cap BD$  being on the cubic, which is equivalent to the triple cross ratio at  $Q$  being  $-1$ .

## 2.7 Exercises/Folklore

1. (folklore) Show that the circular points at infinity are isogonal conjugates with respect to any triangle.
2. (Coharmonic hexagons) This is a generalization of the harmonic quadrilateral.
  - (a) Suppose that  $\angle A + \angle C + \angle E = 360^\circ$  and

$$\frac{AB}{BC} \cdot \frac{CD}{DE} \cdot \frac{EF}{FA} = 1.$$

Show that there exists a point  $M$  such that

$$\triangle MAB \sim \triangle MED,$$

$$\triangle MBC \sim \triangle MFE,$$

$$\triangle MCD \sim \triangle MAF.$$

3. (Isogonal conjugates of a quadrilateral)
  - (a) Show that in a quadrilateral  $ABCD$ , a point  $P$  has an isogonal conjugate (“ $Q$ ”) if and only if  $\angle(AP, PB) + \angle(CP, PD) = 0$  (with directed angles).
  - (b) Show that the locus of such points is a cubic.
  - (c) Show that  $ABPCDQ$  is coharmonic.
  - (d) Show that the midpoints of  $AC, BD, PQ$  are collinear.
  - (e) Prove Fact 10.

## 2.8 Problems

1. (IMO 2019/6) Let  $I$  be the incentre of acute triangle  $ABC$  with  $AB \neq AC$ . The incircle  $\omega$  of  $ABC$  is tangent to sides  $BC, CA$ , and  $AB$  at  $D, E$ , and  $F$ , respectively. The line through  $D$  perpendicular to  $EF$  meets  $\omega$  at  $R$ . Line  $AR$  meets  $\omega$  again at  $P$ . The circumcircles of triangle  $PCE$  and  $PBF$  meet again at  $Q$ .

Prove that lines  $DI$  and  $PQ$  meet on the line through  $A$  perpendicular to  $AI$ .

2. (Romania TST 2017/3/4) Let  $ABCD$  be a convex quadrilateral and let  $P$  and  $Q$  be variable points inside this quadrilateral so that  $\angle APB = \angle CPD = \angle AQB = \angle CQD$ . Prove that the lines  $PQ$  obtained in this way all pass through a fixed point, or they are all parallel.
3. (Sharygin 2024 II/9.3) Let  $(P, P')$  and  $(Q, Q')$  be two pairs of points isogonally conjugated with respect to a triangle  $ABC$ , and  $R$  be the common point of lines  $PQ$  and  $P'Q'$ . Prove that the pedal circles of points  $P, Q$ , and  $R$  are coaxial.
4. (USA TST 2016/3) Let  $ABC$  be an acute scalene triangle and let  $P$  be a point in its interior. Let  $A_1, B_1, C_1$  be projections of  $P$  onto triangle sides  $BC, CA, AB$ , respectively. Find the locus of points  $P$  such that  $AA_1, BB_1, CC_1$  are concurrent and  $\angle PAB + \angle PBC + \angle PCA = 90^\circ$ .
5. (USAMO 2023/6) Let  $ABC$  be a triangle with incenter  $I$  and excenters  $I_a, I_b$ , and  $I_c$  opposite  $A, B$ , and  $C$ , respectively. Let  $D$  be an arbitrary point on the circumcircle of  $\triangle ABC$  that does not lie on any of the lines  $II_a, I_bI_c$ , or  $BC$ . Suppose the circumcircles of  $\triangle DII_a$  and  $\triangle DI_bI_c$  intersect at two distinct points  $D$  and  $F$ . If  $E$  is the intersection of lines  $DF$  and  $BC$ , prove that  $\angle BAD = \angle EAC$ .
6. (239 Olympiad 2010/9.7) In a convex quadrilateral  $ABCD$ , We have  $\angle B = \angle D = 120^\circ$ . Points  $A', B'$  and  $C'$  are symmetric to  $D$  relative to  $BC, CA$  and  $AB$ , respectively. Prove that lines  $AA', BB'$  and  $CC'$  are concurrent.
7. (ISL 2024 G5) Let  $ABC$  be a triangle with incentre  $I$ , and let  $\Omega$  be the circumcircle of triangle  $BIC$ . Let  $K$  be a point in the interior of segment  $BC$  such that  $\angle BAK < \angle KAC$ . The angle bisector of  $\angle BKA$  intersects  $\Omega$  at points  $W$  and  $X$  such that  $A$  and  $W$  lie on the same side of  $BC$ , and the angle bisector of  $\angle CKA$  intersects  $\Omega$  at points  $Y$  and  $Z$  such that  $A$  and  $Y$  lie on the same side of  $BC$ .

Prove that  $\angle WAY = \angle ZAX$ .

8. (Taiwan TST 2018/2/3) Given six points  $A, B, C, D, E, F$  such that

$$\triangle BCD \stackrel{+}{\sim} \triangle ECA \stackrel{+}{\sim} \triangle BFA$$

and let  $I$  be the incenter of  $\triangle ABC$ . Prove that the circumcenter of  $\triangle AID, \triangle BIE, \triangle CIF$  are collinear.

9. (Probably very hard) If you got to this point: try to prove some of the named cubics properties.

## A Suggested other readings

- *Evan's MOP 2025 talk* - short punchy version of this session. I also stole a bunch of problems from here
  - <https://web.evanchen.cc/handouts/cubics/cubics.pdf>
- My handouts
  - [https://web.stanford.edu/~linkewei/pdf/2019\\_alg\\_geom.pdf](https://web.stanford.edu/~linkewei/pdf/2019_alg_geom.pdf) - Projective Transformations and Spaces, Moving Points method (albeit briefly), Conics, Complex Points, Involutions, Coharmonic Hexagons and Cubic Curves.
  - [https://web.stanford.edu/~linkewei/pdf/2020\\_isogonal.pdf](https://web.stanford.edu/~linkewei/pdf/2020_isogonal.pdf) - Focusing on isogonal conjugates in quadrilaterals and isoptic cubics.
- *notzeb's Cross Ratios*
  - <https://www.notzeb.com/cross.pdf>
- *Li4's Plane Geometry (translated)* - very advanced and detailed book, Ch 9-11 are directly relevant
  - <https://files.catbox.moe/tlxdyd.pdf>
- *MarkBcc168's DIT article*
  - <https://artofproblemsolving.com/community/c6h1509866p8957048>
- *The method of animation*
  - <https://artofproblemsolving.com/community/c6h1952595p13480666> (by zacchro, yayups and anantmudgal09)
  - <https://vixra.org/pdf/2511.0039v1.pdf> - extension by Alex Wang



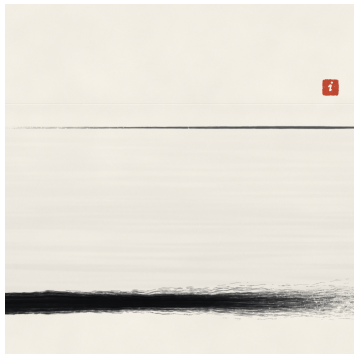
Art by Astra

## B Projective Koans



IN the Zen Buddhist tradition, a *koan* is a riddle given by a master to help their students understand nonduality and experience enlightenment. They appear nonsensical at first, but solving it helps you attain an insight that you cannot learn by reading an explanation.

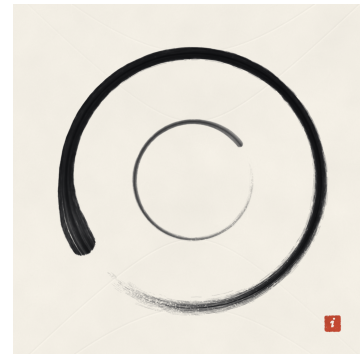
I will borrow the format because I'm too lazy to redo the explanations from my 2019 notes.



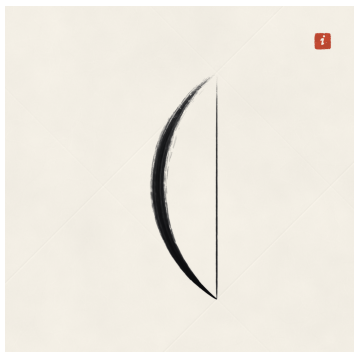
1. Where do two parallel lines intersect?



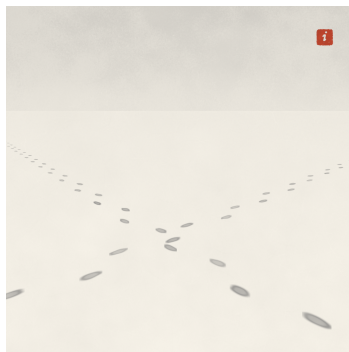
2. Where do a line and a circle intersect, if they do not intersect?



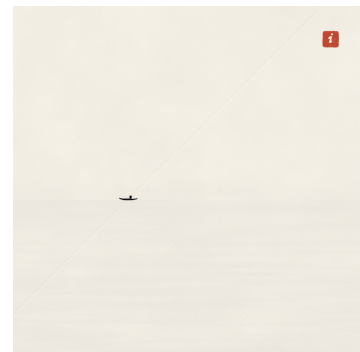
3. Where (and how) do two concentric circles intersect?



4. When are two lines a circle?



5. When can a line be simultaneously parallel to two intersecting lines?



6. What kind of line only has one point on it?



7. Which line is perpendicular to itself?



8. Which angle is bisected by every line through its vertex?



9. Which tangents to a circle pass through its center?

*Koans 6–9 added by Opus 5.5. Images created by Opus 5.5.*

next stop  Cayley–Bacharach →

## C Cayley-Bacharach (but slowly)

Section initially written by me, then rewritten and expanded by Opus 5.5.



HERE'S the puzzle. Five points determine a conic, and by the same kind of count, nine points should determine a cubic. But by Bézout's theorem, two cubics meet in nine points. So which cubic do *those* nine points determine?

### C.1 Pencils and duality

A **conic pencil** is the set of conics passing through four fixed points. You may know about the fact that this set of conics is spanned by any two of them (in the sense that the equation of any third conic you pick is the linear combination of these two).

To demystify this, we'll have to think about what it means for a point to lie on a conic. Let me write down the equation of a conic:

$$A_1x^2 + A_2xy + A_3y^2 + A_4x + A_5y + A_6 = 0.$$

Usually we think of a point on a fixed conic - so as  $x, y$  varies, they satisfy this quadratic constraint. But what about the opposite? If  $x, y$  are fixed and the conic varies, then this is a linear constraint on the coefficients  $(A_1, A_2, A_3, A_4, A_5, A_6)$ . This idea is known as **duality**.

(In my 2019 handout, I present an extremely exciting proof where the main idea is to look at this space of coefficients, which we can aptly call  $\mathbb{P}^5$  - five-dimensional projective space.)

So passing through four points means satisfying four linear equations in six unknowns. If the equations aren't redundant, their solutions form a 2-dimensional space: all the combinations  $\lambda Q_1 + \mu Q_2$  of two particular solutions. That's the pencil. (Five points leave a 1-dimensional space of solutions, which is a single conic, since rescaling an equation doesn't change the curve. That's why five points determine a conic.)

The equations *can* be redundant. If the four points are on a line, then by Bézout every conic through them contains that whole line, so the conics through them are the line together with any other line. That's a 3-dimensional space of equations: only three of the four conditions are doing any work.

### C.2 Nine points, two cubics

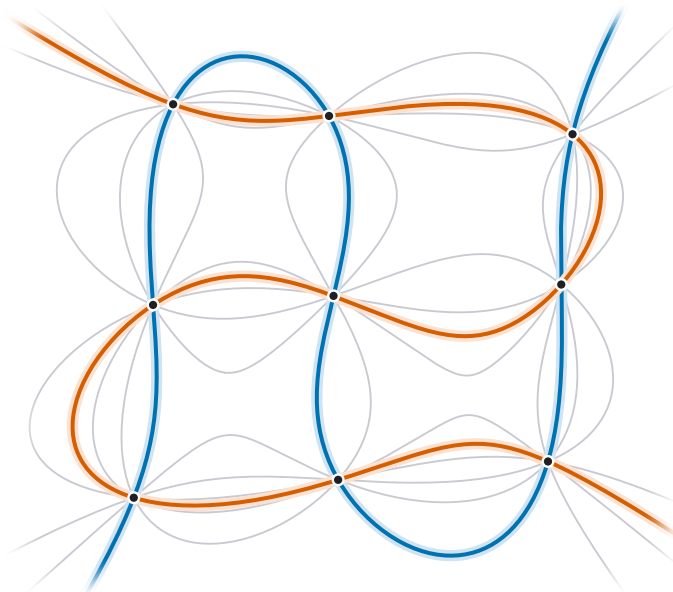
A cubic equation has  $\binom{3+2}{2} = 10$  coefficients, so nine points give nine linear equations in ten unknowns, and you'd expect exactly one cubic through them (up to rescaling). But suppose the nine points are where two cubics  $\mathcal{K}_1 = \{F_1 = 0\}$  and  $\mathcal{K}_2 = \{F_2 = 0\}$  meet. Then every combination  $\lambda F_1 + \mu F_2$  also vanishes at all nine points. The solutions form at least a 2-dimensional space, so the nine equations are only doing the work of eight: one of them is redundant.

That resolves the puzzle: the nine intersection points of two cubics are special, and they don't determine a cubic. Cayley-Bacharach says that the redundancy is spread evenly, so that no particular point is the odd one out.

#### Fact 18 (Cayley-Bacharach)

Suppose  $\mathcal{K}_1$  and  $\mathcal{K}_2$  share no component and meet in nine distinct points. Then the cubics through any eight of the nine points are exactly the pencil  $\lambda F_1 + \mu F_2$ . In particular, every cubic through eight of the points passes through the ninth.

(The second sentence follows from the first, since everything in the pencil passes through the ninth point.)



Two cubics, — and —, meet in nine points, and so does every other cubic in their pencil (grey). The nine points can't single out one cubic.

### C.3 Why eight points are enough

We need the eight points to impose eight *independent* conditions. Then the cubics through them form a  $10 - 8 = 2$ -dimensional space of equations, which has to be the pencil.

First, two observations about our nine points, both from Bézout. **No four of them are on a line**: such a line would meet  $\mathcal{K}_1$  in more than three points, so it would be part of  $\mathcal{K}_1$ , and likewise part of  $\mathcal{K}_2$ . In the same way, **no seven of them are on a conic**. We'll also use the conic fact from above: five points, no four on a line, lie on exactly one conic.

Now suppose the eight points imposed fewer than eight conditions. Then there would be a cubic through them outside the pencil, so the cubic equations through the eight points would form a space of dimension at least 3. Each extra point we demand costs at most one dimension, and here are three ways to spend that freedom, each leading to a contradiction.

- *Three of the points, say  $p_1, p_2, p_3$ , are on a line  $L$ .* Demand one more point of  $L$ . The cubics through all nine points still form a space of dimension at least 2. But each of them meets  $L$  four times, so it is  $L$  together with a conic through the other five points, and there is only one such conic. That's a space of dimension 1.
- *No three are collinear, but six of them, say  $p_1, \dots, p_6$ , are on a conic  $Q$ .* Demand one more point of  $Q$ . Now each cubic meets  $Q$  seven times, so it is  $Q$  together with the line through the remaining two points: again, only one possibility.
- *Neither.* Demand two more points on the line  $L$  through  $p_1$  and  $p_2$ . Some cubic still passes through everything. It meets  $L$  four times, so it is  $L$  together with a conic through  $p_3, \dots, p_8$ . But in this case, no six of the points are on a conic.

So the eight points do impose eight independent conditions, and Cayley–Bacharach follows.

### C.4 Higher degree

The same story works for curves of any degree. Two curves of degrees  $a$  and  $b$  with no common component meet in  $ab$  points, and the general **Cayley–Bacharach theorem** says that a curve of degree  $a + b - 3$  through all but one of those points passes through the last one. For two cubics,  $a + b - 3 = 3$ , which is the statement above.

There's also a refined version, and it is what's behind the quartic rule in Section 1.4. If a new curve of degree  $k$  passes through all but a few of the  $ab$  points, then it is forced through the leftover points too, unless the leftovers fail to impose independent conditions on curves of degree

$a + b - 3 - k$ . For instance, two quartics meet in 16 points. A third quartic through 13 of them is forced through the other 3, unless those 3 are collinear. Here  $a + b - 3 - k = 1$ , and three points fail to impose independent conditions on lines exactly when they're on one line. That's the exception in the quartic rule (Fact 8).

As Section 1.3 pointed out, Cayley–Bacharach and the divisor game each imply the other. The divisor-game direction needs the game's two facts, which get proved in Appendix E.

next stop  Divisor calculus  $\longrightarrow$

## D Why is it called “divisor calculus”?

Section written by me, with minor edits by Opus 5.5.



*What advantages does he derive from the system of book-keeping by double entry! It is among the finest inventions of the human mind.*

*You begin by the form, as if it were the matter.*

*My good friend, you do not see how form and matter are in this case one.*

— Goethe, *Wilhelm Meister’s Apprenticeship*, I.10

THIS will be a bit of a long, diffuse explanation, so buckle up.

### D.1 An allegory

Consider rational functions on  $\mathbb{R} \cup \{\infty\} \rightarrow \mathbb{R} \cup \{\infty\}$ .

- $\frac{x-1}{x-2}$  has a root at 1 and a pole at 2.
- $\frac{(x-1)^2}{(x-2)(x-3)}$  has two roots at 1 and a pole at 2 and 3.
- $x(x-1)$  has a root at 0 and 1. You might think it has no poles, but it has a double pole at infinity - as  $x \rightarrow \infty$ ,  $x(x-1) \sim x^2$ . (In general, if you don’t know the function but know its values, you can tell multiplicity by sampling near a root or pole.)

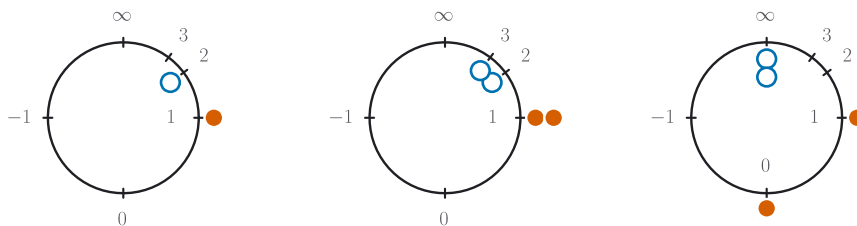
Try as hard as you may, but the number of roots must equal the number of poles.

### D.2 Roots and poles

Let’s say we’re on something like  $\{y^2 = x^3 + 1\}$ . Then, writing something like  $x + 3y$  down gives us a number at each point on the curve. So this is a **function** on the curve.

Now, this thing will have a bunch of roots (where the function is 0) and a bunch of poles. If you had no denominator, you will only get roots, but since we’re working in projective space we’ll get a bunch of poles at the infinite points. A cubic curve typically has 3 points at infinity, so we’ll get  $3d$  poles (where  $d$  is the “leading degree” of the function).

It turns out that the fact from the caricature holds: the number of roots equals to number of poles. This fact follows from Bezout’s theorem: write our rational function as  $A/B$  using homogeneous polynomials of the same degree  $d$ . Then, if the base curve has degree  $e$ , both the numerator and denominator have  $de$  roots with multiplicity, so when netted out we expect the same number of roots and poles.<sup>5</sup>



$\mathbb{R} \cup \{\infty\}$  drawn as a circle, with the roots (•, outside) and the poles (◉, inside) of  $\frac{x-1}{x-2}$ ,  $\frac{(x-1)^2}{(x-2)(x-3)}$  and  $x(x-1)$ . A double root or pole gets two beads.

<sup>5</sup>this also works in the reverse direction: we could also use this to show Bezout’s theorem. Suppose we wanted to show that a cubic  $\{P_3 = 0\}$  and a quintic  $\{P_5 = 0\}$  intersected at 15 points. It suffices to show that  $P_5$  has 15 roots on the cubic. But it does, because it has a quintuple pole at each of the three infinite poles of the cubic.

### D.3 Divisors

With this in mind: a **divisor** is simply a “root accounting” device: positive for roots, negative for poles. It should be intuitive that we can add and subtract divisors. Let’s call the set of divisors  $\text{Div}(C)$ .<sup>6</sup>

A **principal divisor** is one that describes the poles and roots of some rational function of  $x$  and  $y$ . We can also add and subtract these guys (corresponding to multiplying and dividing rational functions). These guys have the same number of roots as poles, so one could say that they have “degree” 0. Not all degree 0 divisors are principal though, we’ll see why in a bit: Appendix E finds the first example, and Appendix F says exactly which ones are. We’ll call this set  $\text{PDiv}(C)$ .

### D.4 Back to the divisor game

The divisor game describes the following set:

$$\text{DivCalc}(C) := \text{Div}(C) / \langle \text{lines} \rangle$$

which spelt out in full means “ $\text{Div}(C)$ , but set the sum of points on each line to 0”. The two facts (Facts 3 and 4) can be rewritten as:

- $\text{Div}(C) / \langle \text{lines} \rangle \cong \text{Div}(C) / \langle \text{algebraic curves} \rangle$
- $[P] - [Q]$  is not principal unless  $P = Q$ .

Conceptually,  $\text{Div}(C) / \langle \text{algebraic curves} \rangle$  is the correct object we’re reasoning about. If we look at the degree 0 piece of this we’ll see why it’s the correct object: the degree 0 part of the “span” of algebraic curves divisors precisely gives us rational functions (a.k.a. principal divisors).

### D.5 Splitting any curve into lines

The first property is actually saying that any polynomial can be written as a product and quotient of linear functions, modulo the base curve. This sounds nontrivial, but surprisingly it boils down to something combinatorial where we’re repeatedly trading in e.g. two positive points for a negative point. Eventually we’re forced to end up with one positive and one negative point, so if we had the second property we’d be done.

### D.6 Picard group

Again, there is standard terminology that refers to something slightly different. The **Picard group** is defined as

$$\text{Pic}(C) := \text{Div}(C) / \text{PDiv}(C),$$

where  $\text{PDiv}(C)$  is the subgroup of principal divisors.

The difference between this group and  $\text{DivCalc}$  is that  $\text{PDiv}(C)$  is the degree 0 piece of the “span” of algebraic curves divisors (which maybe one could reasonably write as  $\langle \text{algebraic curves} \rangle^0$ ), so we have some “leftovers” coming from the degree. For instance, in the Picard group there is exactly one divisor with degree 3, which we can reasonably call the **hyperplane divisor** (usually denoted  $H$ ), and multiples of it are the unique divisors of degree  $3d$ .

One could also write

$$\text{DivCalc}(C) \cong \text{Pic}(C) / \langle H \rangle$$

so you could do all your computations in the Picard group instead of the usual divisor calculus - you’d just get a bunch of  $H$ ’s to deal with as well.

<sup>6</sup>I say “set” here, but really it’s an abelian group.

## D.7 The group law, again

The group law says that there is one object that is canonically identified with points - it's the **Jacobian**:

$$\text{Jac}(C) := \text{Pic}^0(C) = \text{DivCalc}^0(C).$$

Here we use the superscript 0 to denote taking the degree 0 piece.

To map our divisor calculus to the group law, we need a way to map  $\text{DivCalc}$  to  $\text{DivCalc}^0$ . This is actually easy: pick a point  $[O]$  to be zero, then subtract as many copies as you need to get to degree 0.

## D.8 Higher group laws

The general picture has some complications. Roughly speaking,  $\text{Pic}^0(C)$  corresponds (approximately) to the sum of  $g = \frac{(d-1)(d-2)}{2}$  copies of the curve  $C$ , where  $g$  is the **genus** of the curve. In the cubic case,  $g = 1$  and things are straightforward.

For a quartic curve,  $g = 3$ . The rules now read:

- $\text{Div}(C)/\langle \text{lines, conics} \rangle \cong \text{Div}(C)/\langle \text{algebraic curves} \rangle$ .
- $[P_1] + [P_2] + [P_3]$  uniquely identifies the three points  $P_1, P_2, P_3$  (but not their ordering), unless they are collinear. A way to describe this space is that it is mostly the symmetric sum of three copies of  $C$ , and then we need to stitch some collinear triples together (when their respective lines meet at some fixed point).

## D.9 Conclusion

We still have some unanswered questions:

- how do we know that  $P \mapsto [P]$  is injective?
- which degree 0 divisors are principal?
- what does the group law have to do with the torus from Section 2.3?
- where does the genus  $g = \frac{(d-1)(d-2)}{2}$  come from?

The next appendix starts with the first one.

next stop **E** Zooming in  $\longrightarrow$

## E Zooming in

Section written by Opus 5.5, reworking drafts by Astra.



*To see a World in a Grain of Sand  
And a Heaven in a Wild Flower,  
Hold Infinity in the palm of your hand  
And Eternity in an hour.*  
— William Blake, *Auguries of Innocence*

THE previous appendix ended with a question: how do we know that  $P \mapsto [P]$  is injective? We'll see that it comes down to a concrete statement: no rational function on a smooth cubic has just one simple pole. That's a statement about roots and poles, so to attack it we first have to take the root-and-pole accounting seriously.

In the previous appendix, I called a divisor a “root accounting device”, and then cheerfully counted roots and poles “with multiplicity”. But what *is* a double root of a function of  $x$  and  $y$  that lives on a curve?

The answer: if you zoom in far enough, the curve looks like a line, and the function looks like a function of one variable. The allegory from the previous appendix was not really an allegory.

**Ground rules.** From now on, our cubic is **smooth**: no crossings or cusps, including at complex points and at infinity. As usual, we include all complex points and points at infinity. (So the conic-plus-line cubics are benched for this appendix. They come back in the next one.)

Our running example is the same cubic as in the previous appendix,

$$C : y^2 = x^3 + 1.$$

It has a single point at infinity,  $O = [0 : 1 : 0]$ .

### E.1 A ruler at each point

Near  $(0, 1)$ , the curve is the graph of

$$y = \sqrt{1 + x^3} = 1 + \frac{1}{2}x^3 - \frac{1}{8}x^6 + \dots$$

(binomial series), so the single number  $x$  tells us which nearby point we mean. We say that  $t = x$  is a **local coordinate** at  $(0, 1)$ . I like to think of it as a ruler laid along the curve: a function that vanishes at our point and labels the nearby points one-to-one.

Once we have a ruler, every function of  $x$  and  $y$  becomes a function of  $t$  alone. For instance,

$$y - 1 = \frac{1}{2}t^3 - \frac{1}{8}t^6 + \dots = t^3 \left( \frac{1}{2} - \frac{1}{8}t^3 + \dots \right),$$

so  $y - 1$  has a *triple* root at  $(0, 1)$ . Geometrically, the line  $y = 1$  meets the curve three times at  $(0, 1)$  (solve  $x^3 = 0$ ): it is a flex tangent.

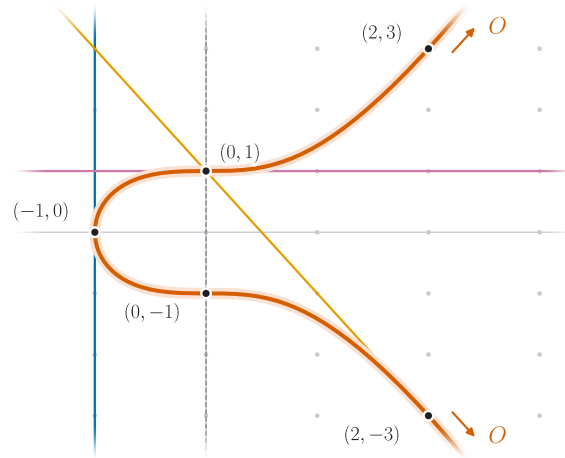
This always works:

#### Fact 19

If  $t$  is a local coordinate at  $P$ , then near  $P$  every nonzero rational function  $f$  can be written as

$$f = t^m u(t),$$

where  $u(t)$  tends to a *nonzero* number as  $t \rightarrow 0$ .



The running example, with the lines whose divisors we compute below:  
 —  $\text{div}(x+1) = 2[(-1, 0)] - 2[O]$ , —  $\text{div}(y-1) = 3[(0, 1)] - 3[O]$ ,  
 —  $\text{div}(y+2x-1) = 2[(2, -3)] + [(0, 1)] - 3[O]$ , —  $\text{div}(x) = [(0, 1)] + [(0, -1)] - 2[O]$ .  
 The last line,  $x = 0$ , passes through the three real flexes  $(0, 1)$ ,  $(0, -1)$  and  $O$ . (The vertical scale is squashed.)

Think of  $u(t)$  as  $c + a_1t + a_2t^2 + \dots$  with  $c \neq 0$ . The integer  $m$  is the **order** of  $f$  at  $P$ , written  $\text{ord}_P(f)$ . If  $m > 0$ , then  $P$  is a root of multiplicity  $m$ ; if  $m < 0$ , it is a pole of order  $-m$ ; if  $m = 0$ , it is neither.

Why is this true?  $f$  is a quotient of two polynomials in  $x$  and  $y$ . Near  $P$ , both  $x$  and  $y$  can be expanded as power series in  $t$  (this is the part we take for granted; it's the implicit function theorem), so the numerator and the denominator become power series in  $t$ . Now factor the lowest power of  $t$  out of each.

So once you zoom in on  $P$ , all that  $f$  remembers is  $t^m$ , exactly as in the allegory. The **divisor** of  $f$  collects these exponents:

$$\text{div}(f) = \sum_P \text{ord}_P(f) [P].$$

Going the other way, you can read a divisor  $\sum n_P [P]$  as a list of instructions: “near each  $P$ , behave like  $t^{n_P}$ ”. A principal divisor is a list of instructions that some actual function obeys. Which lists can be obeyed is the real question, and it will take this appendix and the next to answer it.

Two remarks about rulers:<sup>7</sup>

- They aren't unique. If  $t$  is a ruler, then so is  $2t + t^2$ , or anything of the form  $t \cdot (\text{nonzero} + \dots)$ . Switching rulers doesn't change  $m$ .
- But  $t^2$  is *not* a ruler, since it gives  $t$  and  $-t$  the same label. (So you can't turn a simple root into a double root just by measuring differently.)

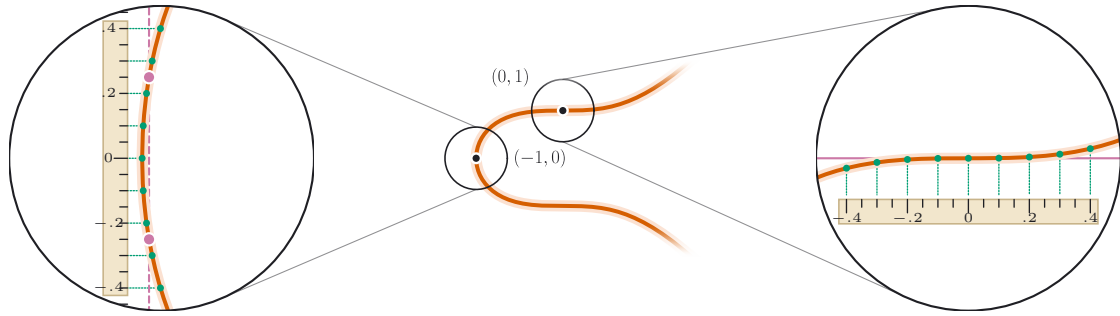
## E.2 When $x$ is the wrong ruler

At  $(-1, 0)$ ,  $x$  stops being a ruler:  $x = -1 + \varepsilon$  corresponds to two nearby points, with  $y \approx \pm\sqrt{3\varepsilon}$ . (The curve has a vertical tangent there.) So we use  $t = y$  instead. Factoring  $x^3 + 1 = (x+1)(x^2 - x + 1)$  gives

$$x + 1 = \frac{y^2}{x^2 - x + 1} = \frac{t^2}{3 + \dots},$$

so  $x + 1$  has a *double* root at  $(-1, 0)$ , even though it is linear! On the curve, it vanishes like the square of the ruler. This is the honest meaning of “a tangent line meets the curve twice at the point of tangency”, which we used all over Section 1.

<sup>7</sup>There are two ways to think about rulers. So far we've used them to measure functions. But they are also gluing instructions: the curve is covered by little patches, each identified with a piece of the complex plane by a ruler, and where two patches overlap, their rulers are related by expansions like  $s = t \cdot (\text{nonzero} + \dots)$ . A space glued together like this is called a **Riemann surface** (a complex manifold of dimension one). It is exactly the setting in which maps between spaces, and integrals along paths, make sense. We'll use both later without needing the name.



Zooming in on the running example. Left: at  $(-1, 0)$ ,  $x$  is the wrong ruler, since  $x = \text{const}$  labels two nearby points at once, but  $y$  gives every nearby point its own label. Right: at  $(0, 1)$ ,  $x$  is a fine ruler, and the flex tangent  $y = 1$  hugs the curve.

### E.3 Zooming in at infinity

At  $O = [0 : 1 : 0]$ , both  $x$  and  $y$  blow up, so neither of them is a ruler. In homogeneous coordinates the curve is  $Y^2Z = X^3 + Z^3$ , and near  $O$  we may set  $Y = 1$ :

$$Z = X^3 + Z^3, \quad \text{so} \quad Z = X^3 + X^9 + \dots$$

Thus  $t = X/Y = x/y$  is a ruler at  $O$ , with  $1/y = Z/Y \approx t^3$  and  $x = X/Z \approx t^{-2}$ . In other words,  $x$  has a double pole at  $O$ , and  $y$  has a triple pole.

Now we can compute some divisors:

$$\begin{aligned} \operatorname{div}(x + 1) &= 2[(-1, 0)] - 2[O], \\ \operatorname{div}(y - 1) &= 3[(0, 1)] - 3[O], \\ \operatorname{div}(y + 2x - 1) &= 2[(2, -3)] + [(0, 1)] - 3[O]. \end{aligned}$$

(The last one is the tangent line at  $(2, -3)$ : substituting  $y = 1 - 2x$  gives  $x(x - 2)^2 = 0$ .) Two things to notice:

- Roots and poles balance every time. Hold that thought.
- If  $\ell = 0$  is a line, written homogeneously, then  $\operatorname{div}(\ell/Z)$  is (the three points where the line meets the cubic) minus  $3[O]$ , because the line at infinity  $Z = 0$  meets the cubic three times at  $O$ . (The previous appendix said that a cubic typically has 3 points at infinity. Ours is atypical, and all 3 pile up at  $O$ .) So the relations of the divisor game, “(points on one line) = (points on another line)”, are exactly the divisors of ratios of two line equations.

### E.4 Following the solutions

Here is a reason that roots and poles balance, which doesn’t use Bézout.<sup>8</sup>

Fix a nonconstant rational function  $f$ . For each number  $w$ , count the solutions of  $f(P) = w$ , with multiplicity (the multiplicity of  $P$  as a root of  $f - w$ ). Call this count  $N(w)$ , and let  $N(\infty)$  be the number of poles.

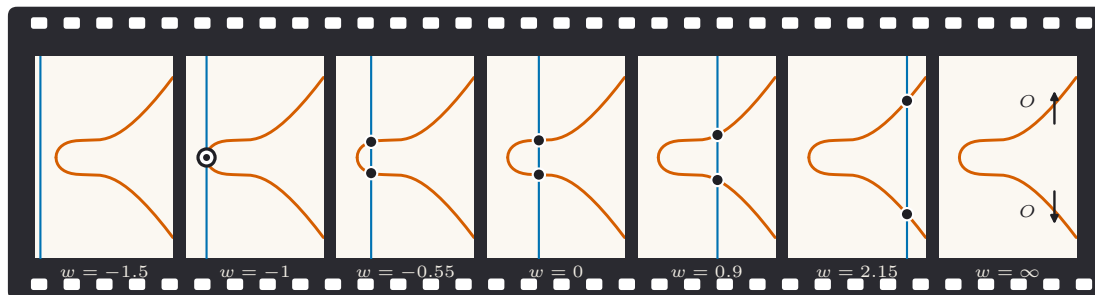
**Claim 20** —  $N(w)$  is the same for every  $w$ , including  $w = \infty$ . In particular,  $N(0) = N(\infty)$ : roots balance poles.

*Zoom in on one solution.* Suppose  $f(P) = w_0$  with multiplicity  $m$ , so  $f - w_0 = t^m u(t)$ . Since  $u$  stays close to  $u(0) \neq 0$ , we can take its  $m$ -th root, and  $s = t \cdot u(t)^{1/m}$  is a new ruler with

$$f - w_0 = s^m.$$

So for  $w$  close to  $w_0$ , the solutions of  $f = w$  near  $P$  are exactly the  $m$  solutions of  $s^m = w - w_0$ . A solution of multiplicity  $m$  is just  $m$  solutions sitting on top of each other: nudge  $w$ , and they split apart. (This is the “sampling near a root” from the allegory.)

<sup>8</sup>Since this argument never used Bézout, footnote 5 in the previous appendix now gives an honest proof of Bézout’s theorem, at least when one of the two curves is our smooth cubic.



The movie for  $f = x$  on  $y^2 = x^3 + 1$ : the solutions of  $\text{—} x = w$ . For  $w < -1$  they are complex, off the picture. At  $w = -1$  they collide, at  $w = 0$  they are the roots  $(0, \pm 1)$ , and as  $w \rightarrow \infty$  they run off to the double pole  $O$ .

*Nothing sneaks in from elsewhere.* Away from the solutions of  $f = w_0$ , the values of  $f$  stay some definite distance away from  $w_0$ . (If not, there would be points where  $f$  gets closer and closer to  $w_0$ . They would pile up somewhere on the curve, and  $f$  would equal  $w_0$  there.) So for  $w$  close enough to  $w_0$ , every solution of  $f = w$  is near a solution of  $f = w_0$ , and  $N(w) = N(w_0)$ .

So  $N$  doesn't change when  $w$  moves a little, which means that it never changes.<sup>9</sup> (Near  $w = \infty$ , run the same argument on  $1/f$  near 0.)  $\square$

I like to picture this as a movie. As  $w$  travels from 0 to  $\infty$ , the solutions of  $f = w$  set out from the roots of  $f$  and arrive at the poles. Along the way they move continuously, sometimes colliding and separating again, but they never appear or disappear.

"They would pile up somewhere on the curve" is where we need *every* point. On the real points alone, solutions can vanish:  $x^2 = w$  has two solutions for  $w > 0$  and none for  $w < 0$  (they've fled to the complex points). On the affine plane alone, solutions can run off to infinity. Including both leaves them nowhere to run. This is the same philosophy as in the projective koans (Appendix B): we are building a world where counting works.<sup>10</sup>

## E.5 Roots and poles pin down a function

Here's a consequence that later appendices use over and over: **a function with no poles is constant.** If  $f$  weren't constant, then  $N(w)$  would equal  $N(\infty) = 0$  for every  $w$ , so  $f$  would take no values at all.

So if two functions  $f$  and  $g$  have the same divisor, then  $f/g$  has no roots and no poles, and  $f = cg$  for a constant  $c$ . A function is determined by its roots and poles up to a constant factor, just as a polynomial is determined by its roots up to the leading coefficient.

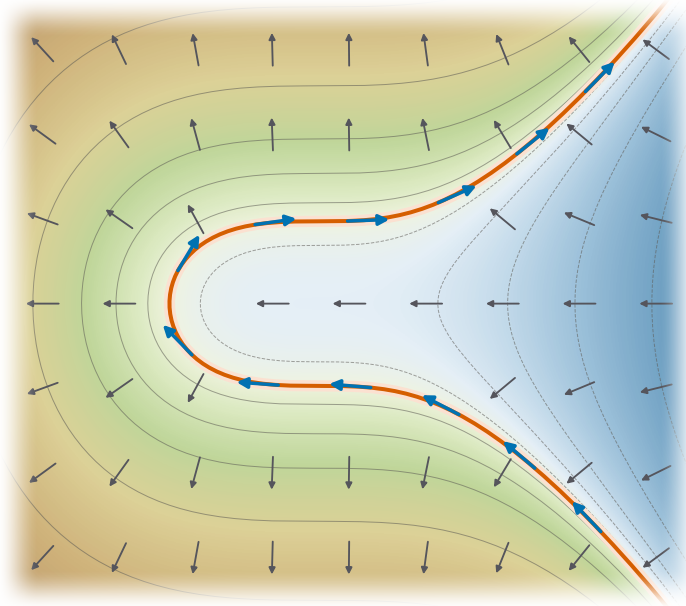
## E.6 A derivative that never breaks

To attack the question from the start of this appendix, we'll want to differentiate functions along the curve, and there's a natural way to do it. Our curve is the zero set of  $F(x, y) = y^2 - x^3 - 1$ . Think of  $F$  as the height of a landscape, so that the curve is the contour line at height 0. The gradient  $(F_x, F_y)$  points straight uphill, perpendicular to the contour line, so turning it by  $90^\circ$  gives an arrow  $(F_y, -F_x)$  that points *along* the curve. Moving along these arrows gives a derivative:

$$Df = F_y \frac{\partial f}{\partial x} - F_x \frac{\partial f}{\partial y} = 2y \frac{df}{dx},$$

<sup>9</sup>Another proof, if you like winding numbers. Walk once around a small loop circling  $P$ . Since  $f \approx ct^m$  there, the value of  $f$  winds  $m$  times around 0. Now cut the whole surface into small tiles. Walking around the edge of a tile,  $f$  winds around 0 as many times as (roots minus poles inside the tile). Add this up over all the tiles: every edge is walked twice, in opposite directions, so the total is 0. (If you meet complex analysis later, this is the argument principle.)

<sup>10</sup>If you've seen some topology: the claim says that  $f$  has a **degree**. A smooth map between closed, oriented surfaces has one: the number of preimages of a point, counted with signs (+1 where the map preserves orientation, -1 where it reverses it), is the same for every point. Here  $f$  goes from the cubic (a closed surface; the next appendix shows it is a torus) to  $\mathbb{C} \cup \{\infty\}$  (a sphere). Complex-differentiable maps never reverse orientation, and near a solution of multiplicity  $m$ ,  $f$  looks like  $s \mapsto s^m$ , which counts as  $m$  preimages. So the signed count is exactly  $N(w)$ .



$F = y^2 - x^3 - 1$  as a landscape, with — the curve as its coastline, at sea level  $F = 0$ . The gradient points uphill (grey arrows); turned by  $90^\circ$ , it blows along the coast (— the wind  $D$ ).

where  $\frac{df}{dx}$  is the derivative as we move along the curve (by implicit differentiation,  $\frac{dy}{dx} = \frac{3x^2}{2y}$ ). For instance,  $Dx = 2y$  and  $Dy = 3x^2$ . The same recipe works for any curve  $F(x, y) = 0$ :  $D = F_y \frac{d}{dx}$ , which is also  $-F_x \frac{d}{dy}$  because  $F_x dx + F_y dy = 0$  along the curve.

Smoothness says exactly that the gradient is never zero, so the arrows never stop at a finite point. The miracle is what happens at infinity. At every point, including  $O$ , in terms of a ruler  $t$  there,

$$D = (\text{something finite and nonzero}) \cdot \frac{d}{dt}.$$

There are three kinds of points to check:

- Where  $y \neq 0$ :  $t = x - x_0$  is a ruler, and  $D = 2y \frac{d}{dt}$  with  $2y \neq 0$ .
- Where  $y = 0$ :  $t = y$  is a ruler, and the chain rule  $\frac{d}{dx} = \frac{dy}{dx} \frac{d}{dy}$  gives  $D = 3x^2 \frac{d}{dt}$  (this is the  $-F_x \frac{d}{dy}$  form). Here  $3x^2 \neq 0$  because  $x^3 = -1$ . (For a general  $y^2 = p(x)$ , this is where we need  $p$  to have distinct roots.)
- At  $O$ :  $t = x/y$ , and  $x \approx t^{-2}$ ,  $y \approx t^{-3}$  give  $Dt = 2y \frac{d}{dt} \approx \frac{2t^{-3}}{-2t^{-3}} = -1$ .

So if  $f$  has order  $m \neq 0$  at  $P$ , then  $Df$  has order exactly  $m - 1$  there (differentiate  $t^m u(t)$ ). If  $f$  is finite at  $P$ , then so is  $Df$ , and  $Df(P) \neq 0$  exactly when  $f - f(P)$  has a *simple* root at  $P$ .

This is special to smooth cubics. (Every smooth cubic has such a derivative, namely  $F_y \frac{d}{dx}$ , although checking it at infinity takes more work in general.) On the projective line, the obvious derivative  $\frac{d}{dt}$  breaks at infinity: in terms of the ruler  $s = 1/t$  there,  $\frac{d}{dt} = -s^2 \frac{d}{ds}$ , which vanishes to second order. In fact, on a line or a conic every derivative of this kind vanishes somewhere, and on smooth curves of degree 4 or more every one of them blows up somewhere. Cubics are the Goldilocks case. (This is the genus from the previous appendix showing its face, and explaining it is where the next appendix begins.)

## E.7 No function has just one simple pole

We can now answer the first question left open at the end of the previous appendix.

**Claim 21** — No rational function on  $C$  has exactly one pole, which is simple.

*Proof.* Suppose  $f$  did, with its pole at  $Q$ .

- At any point  $R \neq Q$ , the function  $f - f(R)$  also has just the one simple pole, so by counting it has exactly one root. That root is  $R$ , so it is simple, and  $Df(R) \neq 0$ .
- Near  $Q$ , we have  $f = c/t + (\text{finite})$ , so  $Df$  has a double pole at  $Q$ .
- Since  $f$  has no other poles, neither does  $Df$ .

So  $Df$  has a double pole and no roots at all. This contradicts roots = poles.  $\square$

On the projective line, the function  $t$  has exactly one simple pole (at  $\infty$ ), so something has to break there. What breaks is that no derivative behaves well everywhere.

Here's another way to see why the claim should be true. If  $f$  had a single simple pole, then by counting it would take every value in  $\mathbb{C} \cup \{\infty\}$  exactly once. So it would perfectly relabel the points of our cubic by the points of  $\mathbb{C} \cup \{\infty\}$ , which is a sphere. In the next appendix we'll see that the cubic is a torus, and you can't continuously relabel a torus as a sphere.

## E.8 Loose ends from Section 1

The claim finishes off both facts that Section 1 stated without proof (Facts 3 and 4).

$P \mapsto [P]$  is **injective**. Suppose  $[P] = [Q]$  in the divisor game. Then  $[P] - [Q]$  is a sum of some line-triples minus some others, and counting points shows that there are equally many of each. So  $[P] - [Q]$  is the divisor of

$$\frac{\ell_1 \ell_2 \cdots \ell_k}{\ell'_1 \ell'_2 \cdots \ell'_k}$$

for some line equations, written homogeneously so that the ratio makes sense at infinity too. Unless  $P = Q$ , this function has a single simple pole.

This also completes the previous appendix's sketch that lines are enough, i.e.  $\text{Div}(C)/\langle \text{lines} \rangle \cong \text{Div}(C)/\langle \text{algebraic curves} \rangle$ : the missing ingredient was injectivity.

**If  $3d$  points sum to zero, they lie on a degree  $d$  curve.** A degree  $d$  equation has  $\binom{d+2}{2}$  coefficients. The equations that are multiples of the cubic's equation vanish on the whole cubic, so they're useless here. Throwing them out leaves

$$\binom{d+2}{2} - \binom{d-1}{2} = 3d$$

coefficients' worth of freedom, so we can find a degree  $d$  curve through any  $3d - 1$  of the points that doesn't contain the cubic. By Bézout, it meets the cubic at one more point  $R$ . Its  $3d$  intersection points sum to zero (lines are enough!), so  $[R]$  equals the last point, and by injectivity,  $R$  is the last point.

So the local picture paid off: near every point, a function is literally a function of one variable, and that was enough to prove both facts. But it raises a bigger question. Each ruler only works near its own point. **Is there one variable that works everywhere?** If there were, say  $z$ , then  $D$  would presumably be the derivative with respect to it:  $D = \frac{d}{dz}$ , so  $\frac{dx}{dz} = Dx = 2y$ , or

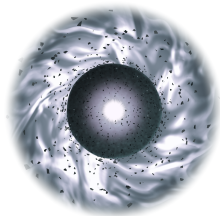
$$dz = \frac{dx}{2y}.$$

This is also the honest version of a question you may have been asking since  $D$  appeared: where does  $D$  come from, and why does it survive at infinity only for cubics? The next appendix answers both.

next stop  Enter the torus  $\longrightarrow$

## F Enter the torus

Section written by Opus 5.5, reworking a first draft by Astra and drawing on my notes.



*I got bored one day and I put everything on a bagel. Everything. All my hopes and dreams, my old report cards, every breed of dog, every last personal ad on Craigslist. . . sesame, poppy seed, salt. . . and it collapsed in on itself. You see, when you put everything on a bagel, it becomes this. The truth. Nothing matters.*

— Jobu Tupaki, *Everything Everywhere All at Once* (2022)<sup>11</sup>

UNFORTUNATELY, the rest of this stuff is going to be incredibly advanced. The good news is that the central calculation isn't, and I think it's worth seeing.

The previous appendix left us with a question: is there one variable that works on the whole curve, the way a ruler  $t$  works near one point? This appendix finds it. The answer explains three things at once: where the derivative  $D$  comes from; why Section 2.3 could say that the degree 0 divisors “actually form a continuous group isomorphic to the torus  $\mathbb{C}/(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$ ”; and why the divisor game's addition is real addition. In the end,  $P \mapsto [P]$  will be a logarithm, quite literally.

As in Section 1, there are two facts that will be stated without (full) proof. I'll flag them as they come.

### F.1 Warm-up: angle chasing is taking logarithms

Let's start with a cubic we've used many times:  $\Gamma \cup \ell_\infty$ , where  $\Gamma$  is the unit circle and  $\ell_\infty$  is the line at infinity. Write points on  $\Gamma$  as complex numbers  $a, b, c, d$  of absolute value 1. Then

$$[A] + [B] = [C] + [D] \iff AB \text{ and } CD \text{ meet } \ell_\infty \text{ at the same point} \iff AB \parallel CD \iff ab = cd.$$

(The last step is a standard complex bash fact: the direction of the chord  $AB$  is pinned down by  $\frac{a-b}{a+b} = -ab$ .) Taking logarithms,

$$\log a + \log b = \log c + \log d \pmod{2\pi i}.$$

So on this cubic,  $[A]$  behaves exactly like  $\log a$ , or, writing  $a = e^{i\theta}$ , like the angle  $\theta$ . **The divisor game on a circle-plus-line is angle chasing.** (Section 1 opened by calling angle chasing a calculus. It turns out to be a special case of ours.)

Where did the logarithm come from? Let's try the recipe from the end of the previous appendix on the circle  $F(x, y) = x^2 + y^2 - 1 = 0$ . Again  $F_y = 2y$ , so we measure with  $dz = \frac{dx}{2y}$ . Substituting  $u = x + iy$ , so that  $x = \frac{u+u^{-1}}{2}$  and  $y = \frac{u-u^{-1}}{2i}$ , a short calculation gives

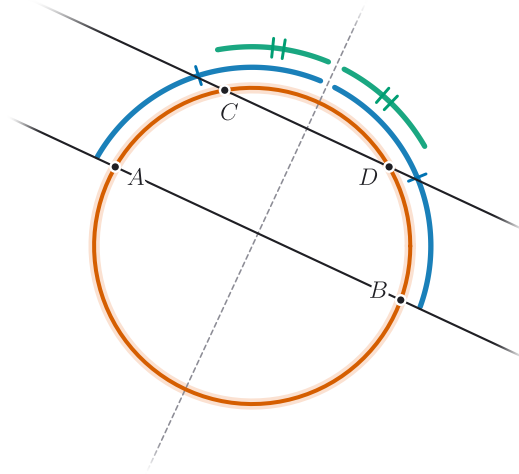
$$\frac{dx}{2y} = \frac{i}{2} \cdot \frac{du}{u}, \quad \text{so} \quad z = \frac{i}{2} \log u.$$

On the real points  $u = e^{i\theta}$ , this is  $z = -\theta/2$ . The measuring rule has rediscovered the angle! So for a circle, the one variable that works everywhere is the angle.

Two more observations:

- $z$  is only defined up to multiples of a single number: going once around the circle changes  $\log u$  by  $2\pi i$ , and hence  $z$  by  $-\pi$ . Such a number is called a **period**.
- The two crossing points of  $\Gamma \cup \ell_\infty$  are the circular points  $J_1 = [1 : i : 0]$  and  $J_2 = [1 : -i : 0]$ . They correspond to  $u = 0$  and  $u = \infty$ , which is exactly where  $\log u$  blows up. No wonder Section 1.2.1 told us to keep away from them.

<sup>11</sup>On screen, the bagel is the donut from Blender's beginner tutorial, with the icing turned off and coloured black (so says the film's VFX artist, Ben Brewer, as reported by 80.lv). Everything really is a torus.



On a circle,  $AB \parallel CD$  exactly when the two chords share an axis of symmetry, that is, when  $\theta_A + \theta_B = \theta_C + \theta_D$ . The ticks mark equal arcs.

A smooth cubic has no crossing points, so its coordinate will never blow up. The price, as we're about to see, is that the curve is a torus, which can be wrapped around in two independent ways.

## F.2 Where does $D$ come from?

Back to the cubic, and to the mystery from the end of Appendix E: why does  $D$  survive at infinity for cubics, and only for cubics?

First, a way to picture  $D$ : it is a **wind** blowing along the surface. At each point it says which way to move and how fast, and  $Df$  measures how quickly  $f$  changes for someone carried along by the wind. (Strictly speaking,  $D$  is a complex derivative, so it describes two winds at right angles to each other, but one is enough for the picture.) “Never breaks” says that the wind never dies down to zero and never becomes infinitely strong.

Since  $D$  is the gradient turned  $90^\circ$ , it can only misbehave where the gradient does, which for a smooth curve means only at infinity. Here's what happens there for a curve of degree  $d$ . Near a typical point at infinity,  $x$  and  $y$  blow up like  $\frac{1}{t}$  in terms of a ruler  $t$ . Then  $F_y$ , which has degree  $d - 1$ , blows up like  $t^{1-d}$ , while  $\frac{dx}{dt} \approx -\frac{1}{t^2}$ . So

$$Dt = F_y \Big/ \frac{dx}{dt} \approx (\text{constant}) \cdot t^{3-d}.$$

For a conic ( $d = 2$ ), the wind stops at each of the two points at infinity. For a cubic, it is finite and nonzero there. (Our  $O$  isn't a typical point at infinity, since all three points at infinity pile up there, but we checked it by hand in Section E.6.) For  $d \geq 4$ , the wind blows up at each of the  $d$  points at infinity. Counting stops as positive and blow-ups as negative, the total is

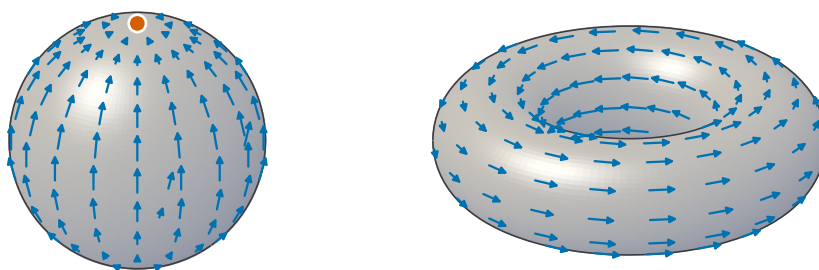
$$d(3 - d).$$

This number is really topology. The **Poincaré–Hopf theorem** says that for any wind on a closed surface with  $g$  holes, the stops minus the blow-ups (each counted with its multiplicity) always add up to  $2 - 2g$ . So

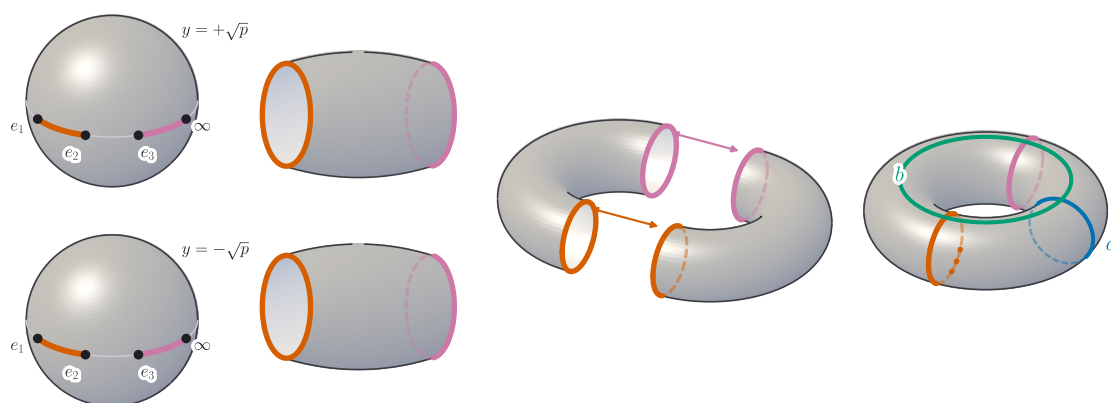
$$2 - 2g = d(3 - d), \quad \text{i.e.} \quad g = \frac{(d-1)(d-2)}{2}.$$

That's the genus from Section D.8, and an answer to the “(What's a genus?)” of Section 1.4: it is the number of holes in the surface. A line or a conic has  $g = 0$  (a sphere), a cubic has  $g = 1$  (a torus), and a smooth quartic has  $g = 3$ .

For a sphere, this is the **hairy ball theorem**: you can't comb a hairy ball flat without leaving a cowlick, so any wind on a sphere must stop somewhere. A torus has no such problem, since the wind can just circulate around the donut. So  $D$  never breaks *because* the cubic is a torus, and, turning this around, the existence of  $D$  already tells us that the cubic is a torus. This explains the Goldilocks remark from Section E.6, and it gives another view of “no function has just one simple pole”: such a function would relabel the cubic as a sphere, and a sphere can't carry  $D$ .



You can't comb a hairy ball: a wind on a sphere has to stop somewhere (here, at the poles). On a torus, it can just circulate.



From left to right: two sheets, each with two slits; open the slits into holes; bend the tubes round; glue. The loop  $a$  goes around the tube, and  $b$  goes along it.

### F.3 Seeing the torus

Poincaré–Hopf says that the complex cubic is a torus, but it's more satisfying to see it. Take  $y^2 = p(x)$ , where  $p(x) = (x - e_1)(x - e_2)(x - e_3)$  has distinct roots.

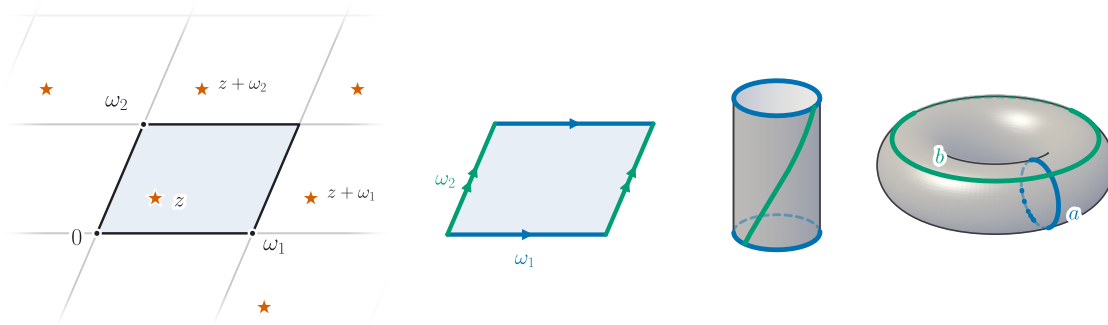
Over each complex number  $x$  there are two points  $(x, \pm\sqrt{p(x)})$ , except over the three roots  $e_1, e_2, e_3$  and over  $x = \infty$ , where there is just one.

Walk around a small loop circling  $e_1$ . Then  $\sqrt{x - e_1}$  changes sign (just like  $\sqrt{x}$  on a loop around 0), while the other two factors come back to where they started. So you come back to the *other* point:  $y$  has become  $-y$ . The same goes for loops around  $e_2$  or  $e_3$ , and for a huge loop around all three (three sign changes make one), which is really a loop around  $\infty$ .

So take two copies of the  $x$ -sphere  $\mathbb{C} \cup \{\infty\}$ , one for each sign of  $y$ , and cut two slits in each: one from  $e_1$  to  $e_2$ , and one from  $e_3$  to  $\infty$ . A loop that doesn't cross a slit circles an even number of the four special points, so the sign of  $y$  is unambiguous; crossing a slit takes you to the other copy. Now open up each slit into a hole. Each copy becomes a sphere with two holes, which is a tube, and crossing a slit glues one tube to the other, end to end. Two tubes glued end to end make a torus!

The real points that we usually draw are just one or two loops on this torus. (For  $y^2 = x^3 + 1$ , the real points form a single loop through  $O$ . When  $e_1, e_2, e_3$  are real, as in the picture, the real points lie exactly over the slits, so they are the two seams: the oval and the unbounded branch.)

Two loops on the torus will matter: one that goes around the tube (say, circling the slit from  $e_1$  to  $e_2$ ), and one that goes along it (crossing into the other copy through one slit and coming back through the other).



Left: in  $\mathbb{C}/\Lambda$ , points that differ by a lattice vector (the stars) are the same point. Right: glue the  $\omega_2$  sides of the parallelogram to get a cylinder, then glue its two ends to get the torus. — The  $\omega_1$  sides become the loop  $a$ , and — the  $\omega_2$  sides become  $b$ .

## F.4 Laying down coordinates

Now we can answer the question from the end of Appendix E. Let the wind carry a point from  $O$  to  $P$ , and keep track of how long the trip takes. That travel time is our coordinate  $z(P)$ , and by construction,  $D$  is just  $\frac{d}{dz}$ . Since  $Dx = 2y$ , i.e.  $\frac{dx}{dz} = 2y$ , the coordinate changes by  $\frac{\Delta x}{2y}$  when  $x$  changes by a tiny  $\Delta x$ . So to compute  $z(P)$ , walk from  $O$  to  $P$  along any path, chop the path into tiny steps, add up  $\frac{\Delta x}{2y}$  over the steps, and take the limit as the steps get smaller:

$$z(P) = \int_O^P \frac{dx}{2y}.$$

The integral sign is just bookkeeping for “add up all the tiny changes”. Since  $x$  and  $y$  are complex, so is  $z(P)$ .

Because the wind never breaks, neither does  $z$ : in terms of any ruler  $t$ ,  $dz = (\text{finite, nonzero}) \cdot dt$ . (Near  $(e, 0)$  it is  $dz = \frac{dy}{3x^2}$ , and near  $O$  it is  $dz \approx -dt$ .) In other words,  **$z$  is itself a valid ruler at every point**: this is the one variable we were looking for. The whole curve is made of small patches that look exactly like pieces of the complex plane. On the overlaps, the local versions of  $z$  differ only by adding constants, so the patches are glued together by sliding, and on each patch the wind blows in a straight line at constant speed. (In the language of footnote 7: this surface is *flat*.)

There’s one catch: does  $z(P)$  depend on the path?

### Fact 22 (Cauchy)

Deforming a path while keeping its endpoints fixed doesn’t change the integral.

(Why you might believe it: zoomed in,  $dz = g(t)dt$  for some power series  $g$ , which has an antiderivative  $G$ . So for paths inside a small patch, the integral is just  $G(\text{end}) - G(\text{start})$ , whatever the path. A big deformation is made of lots of small ones.)

On a torus, though, some loops can’t be shrunk away. Let  $\omega_1$  and  $\omega_2$  be the integrals around the two loops from the end of the last section; these are the **periods**. Any loop can be deformed into “ $m$  times around the first loop, then  $n$  times around the second”, so the integral around it is  $m\omega_1 + n\omega_2$ . So  $z(P)$  is well-defined, as long as we don’t distinguish numbers that differ by an element of the **lattice**

$$\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}.$$

This is what  $\mathbb{C}/\Lambda$  from Section 2.3 means: complex numbers, where two of them count as the same if they differ by an element of  $\Lambda$ . It’s like a clock, where 15 o’clock is 3 o’clock, except that it wraps around in two directions. If  $\omega_1$  and  $\omega_2$  aren’t parallel, every complex number is equivalent to one in the parallelogram with sides  $\omega_1$  and  $\omega_2$ , and gluing opposite sides of the parallelogram gives a torus again.

The **Abel map** is

$$A : C \rightarrow \mathbb{C}/\Lambda, \quad A(P) = \int_O^P \frac{dx}{2y} \pmod{\Lambda}.$$

In particular,  $A(O) = 0$ . (On the circle, the same recipe gave  $\frac{i}{2} \log u$ , with one period.)

## F.5 Collinear points add to zero

Here's the main calculation. A line  $y = mx + c$  meets our cubic at three points  $P_1, P_2, P_3$ , whose  $x$ -coordinates are the roots of

$$q(x) = p(x) - (mx + c)^2 = (x - x_1)(x - x_2)(x - x_3).$$

Keep  $m$  fixed and slide the line by changing  $c$ . The three points move, and I claim that

$$z(P_1) + z(P_2) + z(P_3)$$

doesn't change. (Here we let each  $z(P_i)$  move continuously as an actual complex number, rather than wrapping it back into the parallelogram.)

Differentiating  $p(x_i) = (mx_i + c)^2$  with respect to  $c$  gives

$$p'(x_i) \frac{dx_i}{dc} = 2(mx_i + c) \left( m \frac{dx_i}{dc} + 1 \right), \quad \text{i.e.} \quad \underbrace{(p'(x_i) - 2m(mx_i + c))}_{= q'(x_i)} \frac{dx_i}{dc} = 2(mx_i + c) = 2y_i.$$

(If you'd rather avoid derivatives: change  $c$  by a tiny  $\delta$  and  $x_i$  by a tiny  $\varepsilon$ , expand, and keep only the terms that are linear in  $\delta$  and  $\varepsilon$ .) Since  $dz = \frac{dx}{2y}$ ,

$$\frac{dz(P_i)}{dc} = \frac{1}{2y_i} \cdot \frac{dx_i}{dc} = \frac{1}{q'(x_i)}.$$

Now for the punchline:

$$\frac{1}{q'(x_1)} + \frac{1}{q'(x_2)} + \frac{1}{q'(x_3)} = 0.$$

To see this, note that  $q'(x_i) = \prod_{j \neq i} (x_i - x_j)$ , so the left side is the  $x^2$ -coefficient of the Lagrange interpolation polynomial through  $(x_1, 1)$ ,  $(x_2, 1)$ ,  $(x_3, 1)$ . But that polynomial is the constant 1!

So the sum of the three coordinates is constant as the line slides. To find the constant, slide the line off to infinity. As  $c \rightarrow \infty$ , the line  $Y = mX + cZ$ , i.e.  $\frac{1}{c}Y - \frac{m}{c}X = Z$ , approaches the line at infinity  $Z = 0$ , which meets the cubic only at  $O$ . So all three points approach  $O$ , and their coordinates approach  $A(O) = 0$ . Therefore:

$$\boxed{A(P) + A(Q) + A(R) = 0 \pmod{\Lambda} \quad \text{whenever } P, Q, R \text{ are collinear.}}$$

Some fine print:

- The calculation needs the three points to be distinct, i.e.  $q'(x_i) \neq 0$ . Since  $c$  is a complex number, we can steer around the finitely many bad values of  $c$  (the tangent lines) with little detours. The three points may trade places along a detour, but that doesn't change their sum.
- Tangent lines follow by taking limits. A vertical line meets the cubic at  $(x_0, \pm y_0)$  and  $O$ . Flipping  $y \mapsto -y$  fixes  $O$  and flips the sign of  $\frac{dx}{2y}$ , so  $A(x_0, -y_0) = -A(x_0, y_0)$ , and the sum is again 0.
- Nothing here needed the shape  $y^2 = p(x)$ : for any cubic  $F(x, y) = 0$ , the same computation works with  $dz = \frac{dx}{F_y}$ . (The constant at the end might not be 0, but a better choice of starting point fixes that.) Try it on the circle from the warm-up: you'll find that  $z(X) + z(Y)$  is the same for all parallel chords  $XY$ .

## F.6 What about higher degree?

The same calculation explains the higher group laws of Section D.8. Take a smooth curve  $F(x, y) = 0$  of degree  $d$ . We saw that  $D$  behaves like  $t^{3-d} \frac{d}{dt}$  near each point at infinity, so the measuring rule  $\frac{dx}{F_y}$  behaves like  $t^{d-3} dt$  there, which vanishes once  $d \geq 4$ . That leaves room: multiplying by a polynomial  $h(x, y)$  of degree at most  $d - 3$ , which blows up at most like  $t^{-(d-3)}$ , still gives a measuring rule that never breaks. There are  $1 + 2 + \cdots + (d - 2) = \frac{(d-1)(d-2)}{2} = g$  monomials  $x^i y^j$  of degree at most  $d - 3$ .

Now redo the sliding computation with the rule  $h \frac{dx}{F_y}$ . The line  $y = mx + c$  meets the curve at  $d$  points, whose  $x$ -coordinates are the roots of  $q(x) = F(x, mx + c)$ , and the same differentiation gives

$$\frac{dz_i}{dc} = -\frac{H(x_i)}{q'(x_i)}, \quad H(x) = h(x, mx + c).$$

The Lagrange trick shows that  $\sum_i \frac{H(x_i)}{q'(x_i)} = 0$  whenever  $\deg H \leq d - 2$ : up to the leading coefficient of  $q$ , it is the  $x^{d-1}$ -coefficient of the polynomial through the points  $(x_i, H(x_i))$ , and that polynomial is  $H$  itself. So each of the  $g$  rules makes collinear  $d$ -tuples add up to a constant. (Notice the two constraints pulling against each other: Lagrange would allow  $\deg h \leq d - 2$ , but never breaking at infinity requires  $\deg h \leq d - 3$ .)

So a curve of degree  $d$  gets  $g$  coordinates at once. The **Abel–Jacobi map** sends each point to a vector in  $\mathbb{C}^g$ , defined up to a lattice with  $2g$  generators, one for each basic loop of a surface with  $g$  holes. This counts the genus a second way: as the number of coordinates, rather than the number of holes.

For a smooth quartic, the rules are  $\frac{dx}{F_y}$ ,  $x \frac{dx}{F_y}$  and  $y \frac{dx}{F_y}$ , and points land in a 3-dimensional torus  $\mathbb{C}^3/\Lambda$ . A single point only has one complex dimension, so it can't fill this up, but a triple  $[P_1] + [P_2] + [P_3]$  can. That's why the quartic game is about triples. The rule that a triple is determined by its sum unless it's collinear says that adding up the three coordinates is one-to-one on triples, except that each family of collinear triples (cut out by the lines through a fourth point of the curve) gets squashed to a single point.

## F.7 The group law is addition

Recall the group law from Section 1.2.5: to add  $P$  and  $Q$ , let  $R$  be the third point on line  $PQ$ ; then  $P +_O Q$  is the third point on line  $OR$ . By the box in Section F.5,

$$A(P) + A(Q) + A(R) = 0, \quad A(O) + A(R) + A(P +_O Q) = 0.$$

Since  $A(O) = 0$ , subtracting gives

$$\boxed{A(P +_O Q) = A(P) + A(Q).}$$

More generally, extend  $A$  to divisors by  $A(\sum n_P [P]) = \sum n_P A(P)$ . Every line-triple goes to 0, so **every equation in the divisor game stays true after applying  $A$** . You may replace each  $[P]$  by the complex number  $A(P)$  and compute modulo  $\Lambda$ . In other words, **the game's formal addition is real addition**, on the torus.

To say more, we need the coordinate to be faithful. This is the second black box:

### Fact 23

$\omega_1$  and  $\omega_2$  are not parallel, and  $A$  is a bijection between the cubic and the torus  $\mathbb{C}/\Lambda$ .

Half of this is cheap. Since  $z$  is a valid ruler everywhere, the image of  $A$  can't have an edge, and since the cubic has no missing points, the image must be the whole torus. (The argument from Section E.4 even shows that every point of the torus is hit the same number of times.) The hard part is that this number is 1, i.e. that  $A$  doesn't wrap the cubic around the torus several times. There's a sketch at the end of this appendix, in Section F.10.

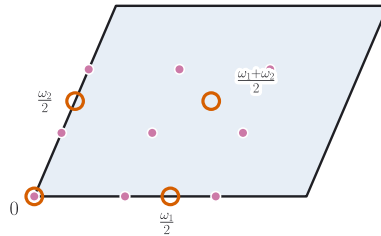
Granting this, the picture is complete.

- **Associativity** of the group law is now free: it's associativity of addition.
- **The degree 0 divisors form the torus.** Every degree 0 divisor can be reduced, using lines, to  $[T] - [O]$  for a single point  $T$  (the “trading” from Section D.5), and  $A$  matches these up one-to-one with  $\mathbb{C}/\Lambda$ . In the language of Appendix D,

$$\text{Jac}(C) = \text{Pic}^0(C) = \text{DivCalc}^0(C) \cong \mathbb{C}/\Lambda.$$

(The full divisor game also remembers the number of points mod 3, which is the invariant we noticed in Section 1.2.4.)

- **Torsion.** The  $m$ -torsion points of  $\mathbb{C}/\Lambda$  form the grid  $\frac{r\omega_1 + s\omega_2}{m}$  for  $0 \leq r, s < m$ , which is the  $(\mathbb{Z}/m\mathbb{Z})^2$  promised in Fact 12. For  $m = 2$ , the nonzero ones are the half-periods  $\frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1 + \omega_2}{2}$ ; on  $y^2 = p(x)$  they are the three points  $(e_i, 0)$  with vertical tangents. The Klein four configurations from Section 2.3 are four points whose coordinates differ by half-periods. For  $m = 3$ , the nine 3-torsion points are the nine flexes.



The 2-torsion points (circles) and the 3-torsion points (dots).

- **Abel's theorem.** Appendix D promised that not every degree 0 divisor is principal, and Appendix E found the first example,  $[P] - [Q]$ . Here is exactly which ones are:

$$\sum_P n_P [P] \text{ is the divisor of a function} \iff \sum_P n_P = 0 \text{ and } \sum_P n_P A(P) = 0.$$

( $\Rightarrow$ : principal divisors are made of lines, since lines are enough, and  $A$  kills lines.  $\Leftarrow$ : reduce to  $[T] - [O]$  using ratios of lines; then  $A(T) = 0$  forces  $T = O$ .)

## F.8 A check you can do on a calculator

Let's test all this on  $y^2 = x^3 + 1$ , which has six points with integer coordinates:

$$O, \quad (2, -3), \quad (0, -1), \quad (-1, 0), \quad (0, 1), \quad (2, 3).$$

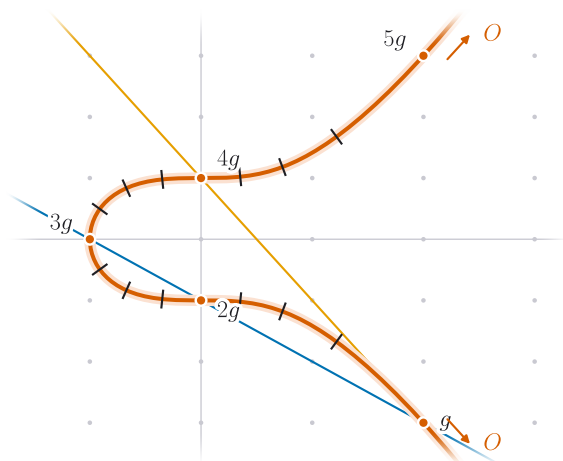
Use the group law with  $O$  as zero (so  $-P$  is the reflection of  $P$  in the  $x$ -axis), and let  $g = (2, -3)$ .

- The tangent at  $g$  is  $y = 1 - 2x$ , which meets the cubic again at  $(0, 1)$ . So  $2g = -(0, 1) = (0, -1)$ .
- The line through  $g$  and  $2g$  is  $y = -1 - x$ , which meets the cubic again at  $(-1, 0)$ . So  $3g = -(-1, 0) = (-1, 0)$ .
- Hence  $4g = -2g = (0, 1)$ ,  $5g = -g = (2, 3)$ , and  $6g = O$ .

Now walk once around the real curve: from  $O$  down the lower branch to  $(-1, 0)$ , then back up the upper branch to  $O$ . Along the way,  $dz = \frac{dx}{2y}$  is always positive (on the lower branch,  $dx$  and  $y$  are both negative), so  $z$  increases steadily from 0 to a real period  $\Omega$ . We pass  $O, g, 2g, 3g, 4g, 5g$  in that order, and  $A$  turns the group law into addition, so these six points must be **evenly spaced**,  $\Omega/6$  apart.

By the symmetry  $y \mapsto -y$ , it's enough to look at the lower branch, where the prediction reads

$$\int_{-1}^0 \frac{dx}{\sqrt{1+x^3}} = \int_0^2 \frac{dx}{\sqrt{1+x^3}} = \int_2^\infty \frac{dx}{\sqrt{1+x^3}}.$$

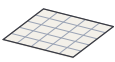


Tick marks at equal steps of  $z$ , four to each gap: the multiples of  $g$  land on every fourth tick. Near  $O$  the ticks spread out, and the first two after  $O$  are off the page.

Three integrals with no apparent reason to agree. Ask your favourite calculator: all three are  $1.40218\dots$ <sup>12</sup>

## F.9 So it really is a logarithm

Let's line up three kinds of "cubic" from this handout. In each case, the measuring rule is  $dz = \frac{dx}{F_y}$ , where  $F(x, y) = 0$  is the curve.

|                      | $y = x^3$                                                                                 | circle + line at infinity                                                                    | smooth cubic                                                                                |
|----------------------|-------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| measuring rule $dz$  | $dx$                                                                                      | $\frac{dx}{2y} = \frac{i}{2} \cdot \frac{du}{u}$                                             | $\frac{dx}{2y}$                                                                             |
| coordinate $z$       | $x$                                                                                       | $\frac{i}{2} \log u$                                                                         | $A(P)$                                                                                      |
| periods              | none                                                                                      | one                                                                                          | two                                                                                         |
| shape                |  plane |  cylinder |  torus |
| collinearity becomes | $x_1 + x_2 + x_3 = 0$                                                                     | $ab = cd$ (parallel chords)                                                                  | $A(P) + A(Q) + A(R) = 0$                                                                    |
| inverse functions    | polynomials                                                                               | cos and sin                                                                                  | $\wp$ and $\wp'$                                                                            |

(The curve  $y = x^3$  has a cusp at infinity, and the circle-plus-line has two crossings. Throwing out those bad points is what leaves a plane and a cylinder.)

The first column is the "elementary version" of the USAMO 2014/3 solution (Problem 2 in Section 1): collinear points on the graph of a cubic polynomial have  $x$ -coordinates with a fixed sum, by Vieta. The middle column is the familiar fact that the logarithm turns multiplication into addition,  $\log(ab) = \log a + \log b$ , at the price of only being defined up to multiples of  $2\pi i$ . The last column is the same story with two periods:  $A$  turns the group law into addition, at the price of only being defined up to  $\Lambda$ . For this reason,  $A$  is also called the **elliptic logarithm**. So when the divisor game writes  $[P]$ , it is really writing " $\log P$ ".

This is a general pattern. A crossing pinches a loop of the surface, so the honest genus goes down, but the divisor game doesn't notice: the pinched loop comes back as a "logarithm" direction (a cylinder) instead of a torus direction. The recipe  $\frac{dx}{F_y}$  handles this by itself, since at a crossing it is allowed to blow up like  $\frac{dt}{t}$  on the two branches, and integrating  $\frac{dt}{t}$  gives a logarithm. For example:

- **Circle plus the line at infinity:** two crossings ( $J_1$  and  $J_2$ ), so the game is one logarithm: angle chasing.
- **Two circles** (a degenerate quartic): four crossings (the two ordinary intersection points, plus  $J_1$  and  $J_2$ ), so the game is three logarithms. That's enough room for a divisor calculus on a pair of circles, with no other curve in sight.

<sup>12</sup>A puzzle for the brave: the substitution  $(x, y) \mapsto (\zeta x, y)$ , where  $\zeta^3 = 1$ , maps the cubic to itself and multiplies  $dz$  by  $\zeta$ . What does that tell you about the shape of the parallelogram for  $y^2 = x^3 + 1$ ?

- **A bicircular quartic** (a quartic with crossings at  $J_1$  and  $J_2$ , like a Cassini oval  $PA \cdot PB = \text{const}$ ): the game is one torus direction plus two logarithms. (The lemniscate has one more crossing, at the origin, and no torus direction left.)

So the game on a degree  $d$  curve always has  $\frac{(d-1)(d-2)}{2}$  dimensions, whatever the singularities. That's why it kept working on circle-plus-line throughout the handout, as long as we stayed away from the crossings.

The inverse of the elliptic logarithm,  $z \mapsto (\wp(z), \frac{1}{2}\wp'(z))$ , plays the role of the exponential: it is the elliptic cousin of  $\theta \mapsto (\cos \theta, \sin \theta)$ . The angle addition formulas for  $\cos$  and  $\sin$  are the group law of the circle, and the addition formula for  $\wp$  is the group law of the cubic. For  $y^2 = x^3 + ax + b$ , it is just Vieta: the  $x$ -coordinates of the three points on  $y = mx + c$  add up to  $m^2$ , so

$$\wp(u) + \wp(v) + \wp(u+v) = \left( \frac{\wp'(u) - \wp'(v)}{2(\wp(u) - \wp(v))} \right)^2.$$

What about the divisor of a function? It's a logarithm too:  $\text{div}(fg) = \text{div}(f) + \text{div}(g)$ , since multiplying functions adds their exponents  $m$ . It even remembers  $\log f$ . Walking once around a small loop circling  $P$ , the value of  $f \approx ct^m$  winds  $m$  times around 0, so  $\log f$  increases by  $2\pi i \cdot \text{ord}_P(f)$ : the divisor records how  $\log f$  fails to be single-valued.

So there are two logarithms in this story, and they fit together as two arrows:

$$\text{functions} \xrightarrow{\text{div}} \text{divisors} \xrightarrow{z} \mathbb{C}/\Lambda.$$

The first turns products of functions into sums of points; the second turns the game's sums of points into honest sums of numbers. Abel's theorem says that the two arrows compose to zero: for every function  $f$ , the roots and poles balance in coordinates as well as in number,

$$\sum_P \text{ord}_P(f) A(P) = 0,$$

and conversely, every degree 0 divisor that balances in this way comes from a function.

So the slogan is: **the divisor game is to cubics what angle chasing is to circles**. And to answer one of the reactions at the very start of this handout (Section 1.1): yes, we really are adding points together. They live on a torus.

If you go looking for more of this, you'll find it under the name "elliptic curves". The next appendix is a translation guide.

## F.10 Why the coordinate wraps around exactly once (optional)

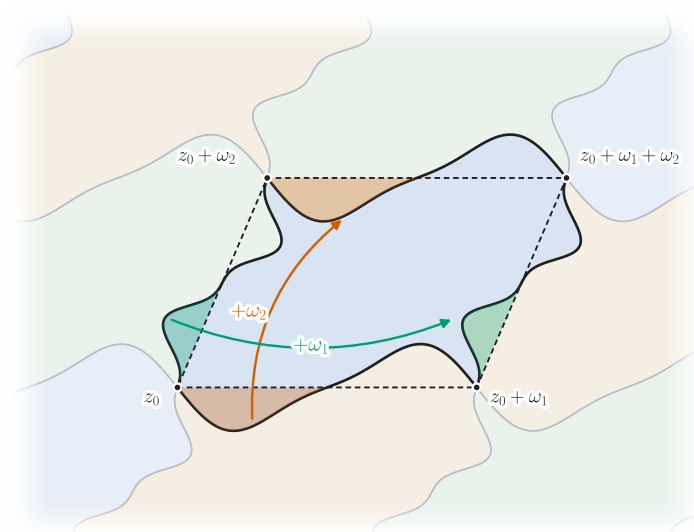
Here is a sketch of the hard half of Fact 23, for those who want it. The idea is to compare areas.

*Cut the torus open.* Cut the cubic along the two loops from before. What's left is a curvy square  $S$  with no holes in it, so  $z$  is single-valued on  $S$  (by Cauchy). Each of the two loops appears twice on the boundary of  $S$ , once on each side of the cut. To get from one copy to the other, you go once around the *other* loop, so the values of  $z$  on the two copies differ by a constant:  $\omega_2$  for one pair, and  $\omega_1$  for the other.

So  $z$  maps the boundary of  $S$  to a closed curve made of four arcs, where opposite arcs are translates of each other by  $\omega_1$  or  $\omega_2$ . It's an Escher-style tile.

*The tile has the same area as the parallelogram.* Each bulge on one side is a dent on the opposite side, so you can cut it off and move it across. So the tile has the same area as the straight parallelogram with sides  $\omega_1$  and  $\omega_2$ . (If you're worried about overlaps, the shoelace formula with signed areas makes this precise.)

*The area of the cubic.* Measure areas on the cubic using  $z$ : a small patch gets the area of its image under  $z$ . By the shoelace formula again (Green's theorem, really), the area of  $z(S)$ , with overlaps counted with multiplicity, is the signed area enclosed by its boundary, which is the area of the tile. Since this area is positive, the parallelogram isn't flat:  $\omega_1$  and  $\omega_2$  are not parallel.



Opposite sides of the tile are translates of each other, so each bulge fills a dent (— move it by  $\omega_2$ , — by  $\omega_1$ ): the tile has the same area as the parallelogram.

*Counting.* On the other hand,  $A$  covers every point of the torus  $\mathbb{C}/\Lambda$  the same number of times, say  $k$ , and it doesn't distort areas (locally, it is just  $z$ ). So the area of the cubic is  $k$  times the area of the parallelogram. Hence  $k = 1$ , i.e.  $A$  is a **bijection**.  $\square$

This is a shadow of Riemann's bilinear relations, if you want to look it up.

next stop  Elliptic curves  $\longrightarrow$

## G Elliptic curves: a dictionary

Section written by Opus 5.5, based on my notes.



*A torsor is like a group that has forgotten its identity.*

— John Baez, *Torsors Made Easy* (2009)

ONE of the reasonable reactions to the IMO 2019/6 solution at the start of Section 1 was “what the heck is an elliptic curve”. Time to answer it.

If you look up “elliptic curve”,<sup>13</sup> you’ll find curves like  $y^2 = x^3 + ax + b$ , along with lots of facts about them. This handout has talked about smooth cubics instead. They are the same thing, and this appendix is a translation guide, so that you can recognize our facts when you meet them in different clothes.

### Fact 24 (Elliptic curves)

An **elliptic curve** is a smooth cubic with a chosen point  $O$  on it. Most references write it in **Weierstrass form**  $y^2 = x^3 + ax + b$ , where  $O$  is the point at infinity  $[0 : 1 : 0]$ .

### G.1 Flexes

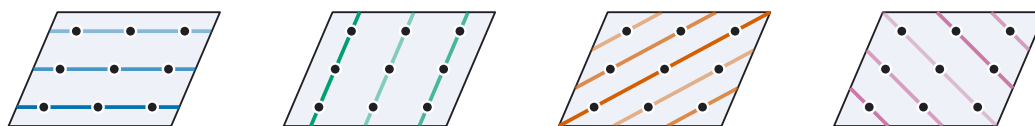
Recall that a **flex** is a point where the tangent line meets the cubic three times. A smooth cubic has exactly 9 of them, and there are two ways to count them.

- **By algebra.** The flexes are the points where the cubic meets its **Hessian**, the curve  $\det(\frac{\partial^2 F}{\partial X_i \partial X_j}) = 0$  built from the second derivatives of the cubic’s equation  $F$ . (A flex is where the curve is straight to second order, so it’s not surprising that second derivatives are involved.) The Hessian is another cubic, so Bézout gives  $3 \times 3 = 9$ .
- **By topology.** Take a flex as  $O$ . The tangent at  $P$  meets the cubic three times at  $P$  exactly when  $3[P] = 0$ , i.e. when  $P$  is 3-torsion, and on the torus the 3-torsion points form a  $3 \times 3$  grid.

The torus also says something that Bézout doesn’t: if  $P$  and  $Q$  are flexes, then so is the third point  $R$  on line  $PQ$ , because  $3R = -3P - 3Q = 0$ . So the line through any two flexes passes through a third! Nine points like this can’t all be real. By the **Sylvester–Gallai theorem**, any finite set of real points, not all on one line, has a line through exactly two of them. In fact, a real cubic always has exactly three real flexes, and they are collinear. (For  $y^2 = x^3 + 1$ , they are  $O$ ,  $(0, 1)$  and  $(0, -1)$ , all on the line  $x = 0$ .)

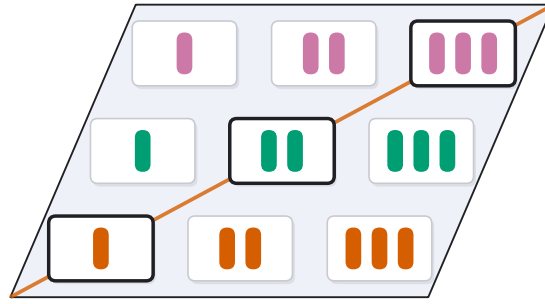
### G.2 Every smooth cubic has a Weierstrass form

The recipe is: find a flex, and send it to infinity. So pick a flex, and use a projective transformation to send it to  $O = [0 : 1 : 0]$  and its tangent line to the line at infinity  $Z = 0$ . That line meets the



The nine flexes as a  $3 \times 3$  grid on the torus. The line through any two flexes passes through a third: the twelve such lines come in four parallel families, wrapping around the torus.

<sup>13</sup>Why “elliptic”? Computing the arc length of an ellipse leads to integrals involving square roots of cubics and quartics, which became known as elliptic integrals. Their inverse functions, like  $\wp$ , are elliptic functions, and the curves on which they live became elliptic curves. Our Abel coordinate  $\int \frac{dx}{2y}$  is the simplest elliptic integral.



The nine flexes play the card game SET. Give the flex in column  $i$  and row  $j$  of the grid ( $i, j = 0, 1, 2$ ) the card with  $i + 1$  ovals in colour  $j$  (red, green, purple). Three cards make a set when each attribute is all the same or all different, which happens exactly when the  $i$ 's and the  $j$ 's each add up to  $0 \bmod 3$ , i.e. when the three flexes are collinear. So any two cards complete to exactly one set, just as the line through two flexes passes through a third. The outlined cards are one set, on one of the — lines of the previous picture.

cubic only at  $O$ , three times, so setting  $Z = 0$  in the cubic's equation leaves just a multiple of  $X^3$ . Thus the equation has the form

$$cX^3 + Z \cdot G(X, Y, Z) = 0, \quad G \text{ quadratic,}$$

and smoothness at  $O$  forces  $G$  to contain a  $Y^2$  term. Setting  $Z = 1$  and dividing by the coefficient of  $y^2$  gives

$$y^2 + (\text{linear in } x) \cdot y = (\text{cubic in } x).$$

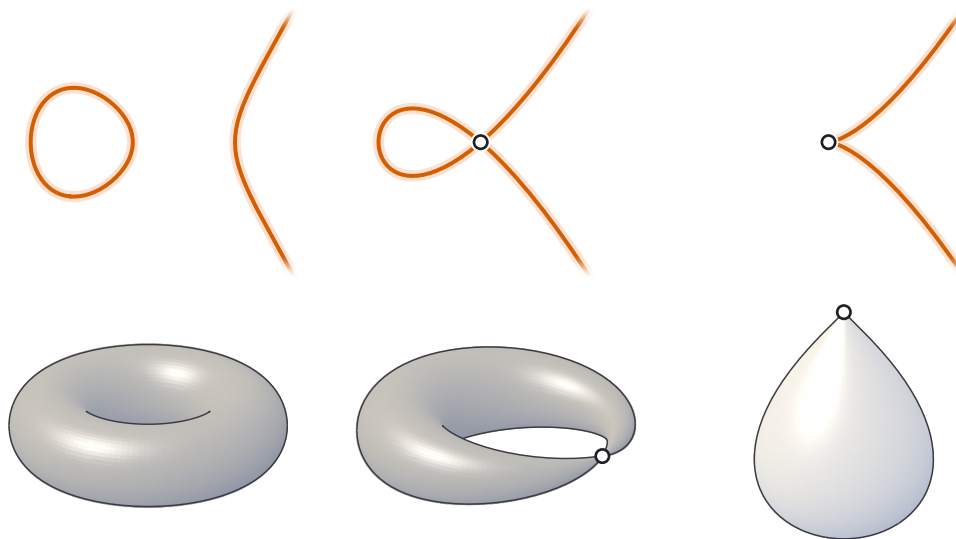
Completing the square in  $y$  removes the middle term, rescaling makes the cubic monic, and shifting  $x$  removes the  $x^2$  term. What's left is  $y^2 = x^3 + ax + b$ , and smoothness says that  $x^3 + ax + b$  has no repeated root.

So nothing is lost by working in Weierstrass form, as long as the chosen point is a flex. (If the chosen point is not a flex, the curve still has a Weierstrass form, but you can't get there by a projective transformation of the plane.)

### G.3 The dictionary

| What you'll read                                                     | What it means in this handout                                                                                                                                                         |
|----------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| elliptic curve $E : y^2 = x^3 + ax + b$ , with $4a^3 + 27b^2 \neq 0$ | a smooth cubic; the condition says $x^3 + ax + b$ has no repeated root                                                                                                                |
| the point at infinity $O$ is the identity                            | a flex, chosen as the zero of $+o$                                                                                                                                                    |
| $P + Q + R = O$ when $P, Q, R$ are collinear                         | $[P] + [Q] + [R] = 0$                                                                                                                                                                 |
| $P + Q$                                                              | $P +_O Q$ , i.e. $[P] + [Q] - [O]$                                                                                                                                                    |
| $-P = (x, -y)$                                                       | the third point on line $OP$                                                                                                                                                          |
| $E[n] \cong (\mathbb{Z}/n\mathbb{Z})^2$ , the $n$ -torsion           | $n$ -torsion degree 0 divisors                                                                                                                                                        |
| $E[2] = \{O, (e_1, 0), (e_2, 0), (e_3, 0)\}$                         | Klein four configurations                                                                                                                                                             |
| $E[3]$                                                               | the 9 flexes                                                                                                                                                                          |
| $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$ , the Abel–Jacobi map       | Appendix F                                                                                                                                                                            |
| the Weierstrass functions $\wp, \wp'$                                | the inverse of the Abel map: $(x, y) = (\wp(z), \frac{1}{2}\wp'(z))$                                                                                                                  |
| divisors, $\text{Pic}^0(E)$ , the Jacobian                           | the divisor game in degree 0                                                                                                                                                          |
| the $j$ -invariant                                                   | the shape of the lattice $\Lambda$ up to rotation and scaling; it is determined by the cross ratio of the four tangents from any point of the cubic (Salmon's theorem, in Appendix I) |

Things you'll find online that this handout doesn't touch: points with rational or integer coordinates (Mordell's theorem, ranks), curves over finite fields (cryptography), and modularity (Fermat's Last Theorem). Those are about the *arithmetic* of the equation; we have only cared about its geometry over the complex numbers.



Top: a smooth cubic, a cubic with a crossing ( $y^2 = x^2(x+1)$ ) and one with a cusp ( $y^2 = x^3$ ). Bottom: their complex points. The crossing pinches one loop of the torus, and the cusp pinches both, which leaves a sphere with a sharp point. Throw out the bad point, and what's left is a cylinder or a plane.

## G.4 Two warnings

- **Choosing a zero.** Online, everything is measured from a fixed point  $O$ . The divisor game of Section 1 is “indecisive” about  $O$ : it works with all points at once, through relations like  $[A] + [B] = [C] + [D]$  that don’t depend on any choice. To translate a statement about  $+_O$  into the divisor game, replace each point  $P$  by  $[P] - [O]$ .
- **Coordinates.** References usually assume Weierstrass form for convenience. Statements that don’t mention coordinates (the group law, torsion, the torus) hold for every smooth cubic, including the circular cubics and isocubics of Section 2. Formulas with explicit  $x$  and  $y$ , like the formula for the coordinates of  $2P$ , only make sense in Weierstrass form.

## G.5 Singular cubics

A cubic with a crossing or a cusp is not an elliptic curve. Its torus degenerates: a crossing pinches one of the two loops, leaving a cylinder whose coordinate is a logarithm (as for the circle-plus-line), and a cusp pinches both, leaving a plane (as for  $y = x^3$ ). These are the first two columns of the table in Section F.9.

next stop  Isocubics  $\rightarrow$

## H Why only two kinds of isocubics?

Section written by Opus 5.5, based on my notes.



SECTION 2.4 quoted a classification theorem: if a cubic is invariant under isogonal conjugation, then conjugation acts on it either as  $X \mapsto P - X$  (a pivotal isocubic) or as  $X \mapsto X + d$  with  $d$  2-torsion (a non-pivotal one). Why only these two? With the torus in hand the answer is short, and along the way it explains why the incenter and excenters, and  $A, B, C$  and the pivot, keep showing up in fours. Throughout,  $z$  is the torus coordinate from Appendix F.

### H.1 Every symmetry is affine

Suppose  $\varphi$  maps the cubic to itself, bijectively, and both  $\varphi$  and its inverse are given by rational functions. Isogonal conjugation, restricted to one of the isocubics of Section 2, is an example.

**Claim 25** — In the torus coordinate,  $\varphi$  is  $z \mapsto az + b$  for some constants  $a$  and  $b$ .

*Proof.* Carry the never-breaking derivative  $D$  across  $\varphi$ : to differentiate a function  $f$  at a point  $Q$ , pull  $f$  back to  $\varphi^{-1}(Q)$ , differentiate there with  $D$ , and come back. In symbols,  $D'f = (D(f \circ \varphi)) \circ \varphi^{-1}$ . Since  $\varphi$  sends rulers to rulers,  $D'$  never breaks either. So in terms of any ruler  $t$ , both  $D$  and  $D'$  are (finite, nonzero)  $\cdot \frac{d}{dt}$ , and their ratio  $D'/D$  is a function with no roots and no poles. By Section E.5, it is a constant  $a$ . Applying  $D' = aD$  to the coordinate  $z$  gives  $D(z \circ \varphi) = a$ , and since  $D = \frac{d}{dz}$ , this says  $z \circ \varphi = az + b$ .  $\square$

### H.2 Involutions: half-turns and half-period shifts

If  $\varphi$  is an involution, then  $a(az + b) + b = z$  on the torus, so  $a^2 = 1$ . There are two cases.

- $a = -1$ :  $\varphi(z) = -z + b$ , a **half-turn** of the torus. It has exactly four fixed points, the solutions of  $2z = b$ . Since  $z + \varphi(z) = b$  is constant, the line through  $X$  and  $\varphi(X)$  always passes through the same point of the cubic, the one with coordinate  $-b$ .
- $a = 1$ :  $\varphi(z) = z + d$  with  $2d = 0$  and  $d \neq 0$ , a **half-period shift**. It has no fixed points, and  $[\varphi(X)] - [X] = d$  is the same 2-torsion element for every  $X$ .

The first case is a pivotal isocubic, with the pivot at coordinate  $-b$ . The four fixed points of isogonal conjugation in a triangle are  $I, I_A, I_B, I_C$ , which is why all four of them double to the same point. The second case is a non-pivotal isocubic, such as the isoptic cubic of a quadrilateral (Section 2.2). So this is the classification theorem quoted in Section 2.4. Notice that  $\varphi$  never had to preserve lines (isogonal conjugation certainly doesn't): being given by rational functions is enough.

Here's a consistency check. A tangential quadrilateral's incenter is its own isogonal conjugate, and it lies on the isoptic cubic. But a half-period shift has no fixed points! The way out is that this cubic isn't smooth: it has a crossing at the incenter.



Pivotal (left): the half-turn  $z \mapsto -z + b$ , which fixes four points,  $I, I_A, I_B, I_C$ . Non-pivotal (right): the shift  $z \mapsto z + \frac{\omega_1}{2}$ , which fixes nothing.

### H.3 Symmetries that preserve lines

Which of the maps  $z \mapsto \pm z + b$  send collinear triples to collinear triples? We need  $(\pm p + b) + (\pm q + b) + (\pm r + b) = 0$  whenever  $p + q + r = 0$ , which means  $3b = 0$ . So on a typical cubic there are 18 of them:

- the translations  $z \mapsto z + \tau$  by the nine 3-torsion points, and
- the half-turns  $z \mapsto -z + \tau$  with  $3\tau = 0$ . These are the **flex projections**: for a flex  $F$  with coordinate  $f$ , the map  $P \mapsto$  (third point on line  $FP$ ) is  $z \mapsto -z - f$ .

One can show that a symmetry of the cubic that preserves collinearity comes from a projective transformation of the whole plane. For example, on  $y^2 = x^3 + 1$ , the flex projection from  $F = (0, 1)$  works out to be

$$(x, y) \mapsto \left( \frac{-2x}{y+1}, \frac{3-y}{y+1} \right),$$

which is a projective transformation. (Check that it maps the cubic to itself!) Composing two flex projections gives a translation by 3-torsion, for the same reason that two half-turns of the plane make a translation.

### H.4 Halving

On a torus, the equation  $nX = c$  has exactly  $n^2$  solutions: if  $X_0$  is one of them, the others are  $X_0$  plus the  $n$ -torsion points. (For  $n = 3$  and  $c = 0$ , these are the nine flexes.) Take  $n = 2$ . The tangent at  $X$  passes through a point  $Q$  of the cubic exactly when  $2X + Q = 0$ , so:

#### Fact 26

From a point  $Q$  on a smooth cubic (not a flex), there are exactly four tangent lines that touch the cubic somewhere else.

The four points of tangency differ from each other by the half-periods  $0, \frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1 + \omega_2}{2}$ , so the sum of any two of them equals the sum of the other two. Translating back into geometry:

#### Fact 27

If the tangents at four points  $A, B, C, D$  of a smooth cubic all pass through one point of the cubic, then the three diagonal points  $AB \cap CD$ ,  $AC \cap BD$  and  $AD \cap BC$  also lie on the cubic.

In a pivotal isocubic with pivot  $X$ , this happens twice. The tangents at  $I, I_A, I_B, I_C$  all pass through  $X$ , and the diagonal points of  $II_A I_B I_C$  are  $A, B$  and  $C$ . The tangents at  $A, B, C, X$  all pass through the conjugate  $X'$ , and the diagonal points are the feet of the cevians through  $X$ . Section 2.4 listed both facts separately; now they are one statement.

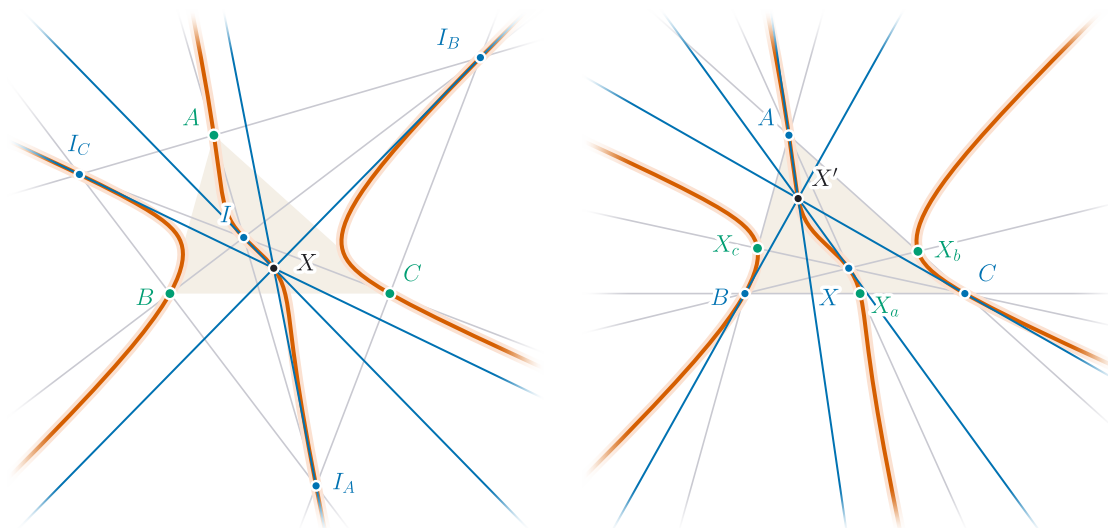
### H.5 Thirds

What if  $A - B = B - C = C - A$ ? Adding up the three differences gives  $3(A - B) = 0$ , so  $\{A, B, C\} = \{X, X + \tau, X + 2\tau\}$  for a 3-torsion element  $\tau$ . That is,  $A, B, C$  is an orbit of one of the translations above, which come from projective transformations of order 3. Geometrically,  $A - B = B - C$  says  $2B = A + C$ : the tangent at  $B$  meets line  $AC$  on the cubic. So:

#### Fact 28

Suppose triangle  $ABC$  is inscribed in a smooth cubic, and the tangents at  $B$  and at  $C$  meet the opposite sides on the cubic. Then so does the tangent at  $A$ . Moreover, some projective transformation that preserves the cubic sends  $A \rightarrow B \rightarrow C \rightarrow A$ .

Through each point of the cubic there are exactly four such triangles, one for each pair  $\pm\tau$  of nonzero 3-torsion elements.

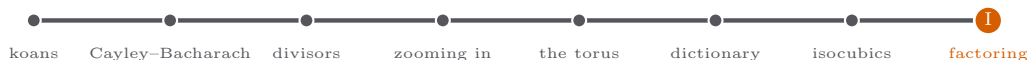


The pivotal isocubic from Section 2.4. Left: the tangents from  $X$  touch at  $I, I_A, I_B, I_C$ , and the diagonal points of these four are  $A, B, C$ . Right: the tangents from  $X'$  touch at  $A, B, C, X$ , and their diagonal points are the feet  $X_a, X_b, X_c$ .

next stop **I** Factoring functions  $\longrightarrow$

# I Factoring functions (optional)

Section written by Opus 5.5, based on my notes.



*Ideal complex numbers are comparable to hypothetical radicals, which do not exist by themselves, but only in combinations; fluorine in particular, an element no one knows how to isolate, can be compared to an ideal prime factor.*

— Ernst Kummer, on ideal numbers (1851)

**D**IVISORS are root-and-pole accounting. Appendix E showed that near any one point, a function really is a power series in one variable, and Appendix F found the variable  $z$  that works everywhere. So can we go all the way, and factor a function over its roots and poles, the way a polynomial factors over its roots?

On a line, every rational function factors into **atoms**:

$$f(x) = c(x - a_1)^{n_1}(x - a_2)^{n_2} \cdots (x - a_k)^{n_k}.$$

Each atom  $x - a$  carries a single root, so a function is just its divisor plus a constant. On a cubic there are no honest atoms (by Appendix E, no function has just a single simple root and a single simple pole). This appendix explains the next best thing. It uses heavier machinery from complex analysis, so treat it as optional.

## I.1 Atoms on a circle

On the unit circle, use the angle  $\theta$  as the coordinate. The chord from angle  $\alpha$  to angle  $\theta$  has length  $2|\sin \frac{\theta - \alpha}{2}|$ , and as a function of  $\theta$ ,  $\sin \frac{\theta - \alpha}{2}$  vanishes exactly once per turn, at  $\theta = \alpha$ . So it is an atom, except for one defect: going once around ( $\theta \mapsto \theta + 2\pi$ ) flips its sign. A ratio with equally many atoms upstairs and downstairs doesn't have this problem, and is an honest function on the circle. Trig bashers use these all the time: a chord ratio like  $\frac{AB \cdot CD \cdot EF}{BC \cdot DE \cdot FA}$  is a ratio of six atoms.

## I.2 Atoms on the torus

The torus has atoms too, with a defect much like the circle's. We'll take their existence on faith:

### Fact 29

There is a function  $\sigma(u)$ , defined for every complex number  $u$ , whose roots are exactly the points of the lattice  $\Lambda$ , all simple. It is periodic up to a simple factor: for every  $W \in \Lambda$ ,

$$\sigma(u + W) = \pm e^{\eta(W)(u + W/2)} \sigma(u),$$

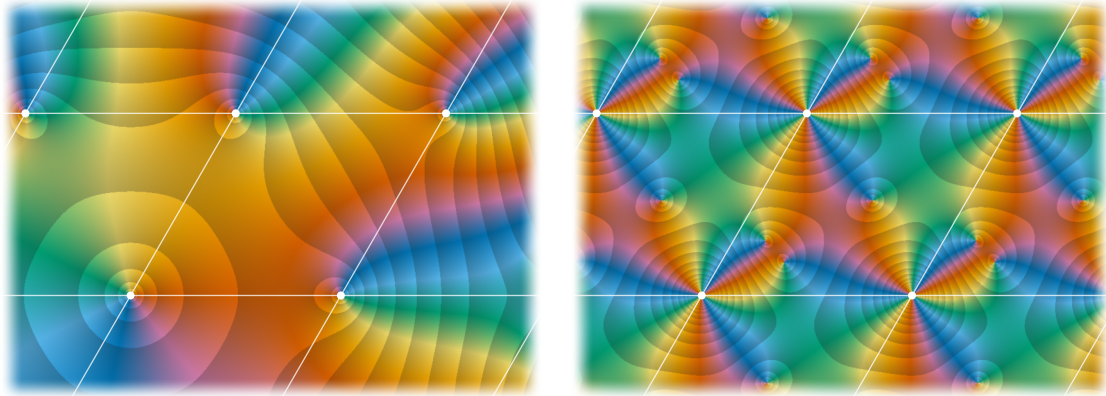
where the sign is  $-$  unless  $\frac{W}{2} \in \Lambda$ , and  $\eta(W + W') = \eta(W) + \eta(W')$ .

This is the **Weierstrass sigma function**. It's built the same way that  $\sin$  can be built as an infinite product over its roots:

$$\sigma(u) = u \prod_{W \neq 0} \left(1 - \frac{u}{W}\right) e^{u/W + u^2/(2W^2)},$$

where the exponential factors are only there to make the product converge.

So  $\sigma(u - a)$  is an atom with a single root at  $u = a$ . The price is the factor  $\pm e^{\eta(W)(u - a + W/2)}$  that it picks up when  $u$  goes around the torus.



Domain colouring: the colour at  $u$  shows the argument of the value, and each band is a doubling of its size.

Left: the atom  $\sigma(u)$  has a simple zero at every lattice point, but no two cells look alike. Right: the balanced product  $\sigma(u-p)\sigma(u-q)\sigma(u-r)/\sigma(u)^3$ , with  $p+q+r=0$ , looks the same in every cell, so it is an honest function on the torus.

### I.3 Balanced products are functions

Take a product of atoms  $f(u) = \prod_i \sigma(u - a_i)^{n_i}$ , listing repeated points separately so that each  $n_i$  is  $\pm 1$ . When  $u$  goes around by  $W$ ,  $f$  gets multiplied by

$$\prod_i \left( \pm e^{\eta(W)(u - a_i + W/2)} \right)^{n_i} = (\pm 1)^{\sum n_i} \exp \left( \eta(W) \left[ \left( u + \frac{W}{2} \right) \sum_i n_i - \sum_i n_i a_i \right] \right).$$

If  $\sum n_i = 0$  and  $\sum n_i a_i = 0$ , this is exactly 1, so  $f$  is an honest function on the torus. These are precisely the conditions in Abel's theorem! Given a degree 0 divisor whose Abel sum is 0 modulo  $\Lambda$ , shift one of the  $a_i$  by a lattice vector to make  $\sum n_i a_i = 0$  exactly, and the product of atoms has precisely the prescribed roots and poles. (One can show that it is a rational function of  $x$  and  $y$ .) This is the “building” half of Abel's theorem, done by hand.

### I.4 Factoring lines

Going the other way, every function factors into atoms, because a function is pinned down by its roots and poles. (The counting argument from Section E.4 works just as well for functions on the torus.) For example, on  $y^2 = x^3 + ax + b$ , suppose a line meets the cubic at the points with coordinates  $p, q, r$ , chosen so that  $p + q + r = 0$  exactly. Then the line's equation, as a function on the cubic, is

$$c \cdot \frac{\sigma(u-p)\sigma(u-q)\sigma(u-r)}{\sigma(u)^3}$$

for some constant  $c$ . Both sides have roots at  $p, q, r$  and a triple pole at  $O$  (where  $u = 0$ ), so their ratio is constant. Similarly, the vertical line through the point with coordinate  $p$  gives the classical formula

$$x - x(p) = - \frac{\sigma(u+p)\sigma(u-p)}{\sigma(u)^2 \sigma(p)^2},$$

where the constant comes from comparing both sides near  $u = 0$ .

### I.5 Determinants

The building block of every cross ratio in this handout is the determinant  $[P, X, Y]$  of the homogeneous coordinates of three points, which vanishes exactly when  $P, X, Y$  are collinear. (The  $[a, b]_p$  of Section 2.6.2 is such a determinant, up to a constant.) With the three points at torus coordinates  $u, v, w$ ,

$$[P, X, Y] = (\text{constant}) \cdot \frac{\sigma(u-v)\sigma(u-w)\sigma(v-w)\sigma(u+v+w)}{\sigma(u)^3 \sigma(v)^3 \sigma(w)^3}.$$

Why? As a function of  $u$  alone, the determinant has a triple pole at  $u = 0$ , and roots exactly when  $P$  is on line  $XY$ , i.e. at  $u = v, w$  and  $-v - w$ . The right side has the same roots and poles, so their ratio doesn't depend on  $u$ . The same goes for  $v$  and for  $w$ , so the ratio is a constant.

Compare this with triangles inscribed in a circle, whose area is  $\frac{abc}{4R}$ : a product of three chords. On the cubic there are three “elliptic chords” and a fourth factor,  $\sigma(u + v + w)$ , which vanishes when the three points are collinear.

## I.6 Salmon’s theorem

From a point  $Q$  of the cubic, there are four tangent lines that touch the cubic somewhere else (see Section H.4). Here is a classical surprise:

### Fact 30 (Salmon’s theorem)

The cross ratio of the four tangents from  $Q$  is the same for every point  $Q$  of the cubic.

Why should it be? Moving  $Q$  moves all four tangents in complicated ways, and there’s no projective transformation in sight. But with atoms it’s a short computation. The cross ratio of four lines through  $Q$  can be written with determinants:

$$(QX_1, QX_2; QX_3, QX_4) = \frac{[Q, X_1, X_3][Q, X_2, X_4]}{[Q, X_2, X_3][Q, X_1, X_4]}.$$

The points of tangency satisfy  $2X_i = -Q$ , so their coordinates are  $x_i = x_0 + h_i$ , where  $2x_0 = -q$  and  $h_1, h_2, h_3, h_4$  are the four half-periods  $0, \frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1 + \omega_2}{2}$ . Plug in the determinant formula. The factors  $\sigma(q)$ ,  $\sigma(x_i)$  and  $\sigma(q - x_i)$  appear equally often upstairs and downstairs, so they cancel. In what’s left,  $x_i - x_j = h_i - h_j$  and  $q + x_i + x_j = h_i + h_j$ , so

$$(QX_1, QX_2; QX_3, QX_4) = \frac{\sigma(h_1 - h_3)\sigma(h_1 + h_3)\sigma(h_2 - h_4)\sigma(h_2 + h_4)}{\sigma(h_2 - h_3)\sigma(h_2 + h_3)\sigma(h_1 - h_4)\sigma(h_1 + h_4)},$$

and  $Q$  has disappeared! To find the actual value, take  $Q = O$ . There the four tangents are the three vertical tangents  $x = e_1, e_2, e_3$  and the line at infinity, so the cross ratio is that of  $e_1, e_2, e_3, \infty$ , in some order. This number is what determines the  $j$ -invariant from Appendix G. For our  $y^2 = x^3 + 1$ , the roots are  $-1$  and  $e^{\pm i\pi/3}$ , and the cross ratio works out to  $e^{i\pi/3}$ : from every point of this cubic, the four tangents form an *equianharmonic* set, the most symmetric kind of quadruple. That’s the same symmetry as the hexagonal lattice in the puzzle of footnote 12.

Here’s a consequence in triangle geometry. Take the pivotal isocubic of triangle  $ABC$  with pivot  $X$  (Section 2.4). The tangents from  $X$  touch it at  $I, I_A, I_B, I_C$ , and the tangents from the isogonal conjugate  $X'$  touch it at  $X, A, B, C$ . So Salmon’s theorem says:

### Fact 31

For any point  $X$  with isogonal conjugate  $X'$ ,

$$(XI, XI_A; XI_B, XI_C) = (X'X, X'A; X'B, X'C).$$

The order matters, and this is the right one. The line  $AI I_A$  meets the cubic at  $A, I, I_A$ , and the tangent at  $I$  meets it at  $I, I, X$ , so  $[I_A] + [A] = [I] + [X]$ . In other words  $I_A - I = X - A$ , which equals  $A - X$  because it is 2-torsion. So the half-period that takes  $I$  to  $I_A$  also takes  $X$  to  $A$ , and similarly for  $B$  and  $C$ .

## I.7 The triple cross ratio

Here is another example. In Section 2.6.2, the triple cross ratio  $T(P) = (A, D; B, E; C, F)_P$  is constant along the cubic through  $A, \dots, F$  and the three diagonal points. (As a function of  $P$ , it is a ratio of line equations whose roots and poles all cancel, so by Section E.5 it is constant.) What is the constant  $k$ ?

Plug the determinant formula into  $T$ , writing  $a, b, \dots, f$  for the coordinates of  $A, \dots, F$ , chosen so that  $a + b = d + e$ ,  $b + c = e + f$  and  $c + d = f + a$  hold exactly. Almost everything cancels, and

$$k = \frac{\sigma(a - b)\sigma(c - d)\sigma(e - f)}{\sigma(b - c)\sigma(d - e)\sigma(f - a)}.$$

This is the elliptic version of the chord ratio  $\frac{AB \cdot CD \cdot EF}{BC \cdot DE \cdot FA}$  for a hexagon on a circle, with  $\sigma$  playing the part of chord length.

Two things fall out of this formula.

- **2-torsion gives  $-1$ .** Let  $\delta = a - d = e - b = c - f$  be the difference element, and substitute  $d = a - \delta$ ,  $e = b + \delta$ ,  $f = c - \delta$ :

$$k = \frac{\sigma(a-b)}{\sigma(a-b-2\delta)} \cdot \frac{\sigma(b-c+2\delta)}{\sigma(b-c)} \cdot \frac{\sigma(c-a+\delta)}{\sigma(c-a-\delta)}.$$

If  $2\delta = W$  is in  $\Lambda$ , each fraction compares  $\sigma$  at two points that differ by  $W$ . The three exponential factors cancel, the three signs multiply to  $-1$ , and  $k = -1$ . (The sketch after Fact 17 is also a complete proof. At  $Q = AE \cap BD$ , the lines  $QA$  and  $QE$  coincide, and so do  $QB$  and  $QD$ , so the triple cross ratio at  $Q$  is  $-1$  for any six points. Since the cubic is the level set  $\{T = k\}$ ,  $Q$  lies on it exactly when  $k = -1$ .)

- **$k$  is not a function of  $\delta$  alone.** With  $B$ ,  $C$  and  $\delta$  fixed, moving  $A$  along the cubic changes  $k$ , unless  $2\delta = 0$ . For example, on  $y^2 = x^3 + 1$  with  $\delta = (2, 3)$  and  $B$ ,  $C$  fixed, three choices of  $A$  give  $k \approx 0.06$ ,  $1.02$  and  $-0.53$ .

## I.8 The allegory, completed

So the allegory from Appendix D has become completely literal. Rational functions on a line are  $c \prod (x-a)^n$ . Functions on a cubic are  $c \prod \sigma(z-a)^n$ , with the torus coordinate  $z$  in place of  $x$  and  $\sigma(z-a)$  in place of  $x-a$ . The only new rule is the one Abel found: the roots and poles have to balance in coordinates as well as in number.

*We shall not cease from exploration  
And the end of all our exploring  
Will be to arrive where we started  
And know the place for the first time.  
— T. S. Eliot, Little Gidding*

*(after initial confusion) we're adding points together now.*

end of the line |



*Moon bridge - by Opus 5.5*