7 Elliptic Curves

To bring the discussion of Fermat's Last Theorem full-circle, we reference another of Fermat's 'margin notes' from his copy of Diophantus' *Arithmetica.* In 1650 Fermat claimed that the equation $y^2 = x^3 - 2$ has only two solutions in integers; namely $(x, y) = (3, \pm 5)$. The first correct proof in writing came around 150 years later.[1] It is perhaps ironic that the proof of Fermat's Last Theorem came via the consideration of *elliptic curves*, of which $y^2 = x^3 - 2$ is an example. Here is the rough timeline:

- (1955-57) Yutaka Taniyama and Goro Shimura conjecture that every elliptic curve over the rational numbers is *modular*. Attempting to define what this means with any precision is beyond this course, though it refers to the *modular group* of integer matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with determinant 1, which play a recurring role in number theory.[2]

- (1984–86) Gerhard Frey, Jean-Pierre Serre and Ken Ribet prove that if (a, b, c) is a non-trivial solution to Fermat's equation $x^p + y^p = z^p$ where p is an odd prime, then the *Frey curve*

  $$y^2 = x(x - a^p)(x + b^p)$$

  is a non-modular elliptic curve.

- (1986–94) Andrew Wiles (and Richard Taylor) prove that all *semistable*[3] elliptic curves are modular. Since the hypothetical Frey curves would be semistable, this shows that they cannot exist and that Fermat's equation therefore has no solutions. Based on the ideas of Wiles and Taylor, the proof of the full Taniyama–Shimura conjecture was eventually completed in 2001 and is now known as the *modularity theorem.*

Wiles' proof is far too difficult for us! In what follows, we discuss some of the beautiful structure of elliptic curves and the way in which their study infuses number theory with geometry and algebra. In particular, we discuss the question of finding integer and rational points on elliptic curves, and some of the *modularity patterns* that arise when considering elliptic curves modulo primes.

Elliptic curves are more than merely interesting to those intent on proving 350-year-old conjectures. They form the basis of a widely-used cryptographic system superior in several ways to the famous RSA system, especially in situations where computing power is at a premium. Their structure also drives the particularly efficient *Lenstra algorithm* for factorizing integers.

Before doing any of this, we need to define an elliptic curve, and then to understand (some) of the components of the definition: this will require a little work...

---

[1] Euler thought he had proved this earlier, but he implicitly assumed unique factorization in the ring $\mathbb{Z}[i\sqrt{2}]$: once this is established the result is straightforward: write $x^3 = (y + i\sqrt{2})(y - i\sqrt{2})$, show that the two factors on the RHS are coprime, whence each must be a perfect cube in $\mathbb{Z}[i\sqrt{2}]$; but if $y + i\sqrt{2} = (a + ib\sqrt{2})^3$, then $(a, b) = \ldots$

[2] Indeed we've met such several times in this course, for instance the matrices of successive convergents $\left(\begin{smallmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{smallmatrix}\right)$ of continued fractions.

[3] As with modularity, a thorough definition of what this means requires a lot more algebraic geometry, though we'll be able to give a loose version at the end of the chapter.

7.1 Algebraic Curves

Definition 7.1. A *planar algebraic curve* $\mathcal{C}$ is the set of points (x, y) satisfying some polynomial equation $f(x, y) = 0$. The *degree* of $\mathcal{C}$ is the degree of the polynomial f.

A curve is *irreducible* if its defining polynomial is irreducible:

$$f(x, y) = g(x, y)h(x, y) \implies g \text{ or } h \text{ is constant}$$

The values x, y can lie in various fields. Unless otherwise stated, we assume $x, y \in \mathbb{R}$, though sometimes $\mathbb{Q}$ or $\mathbb{C}$ are more appropriate. Later in the chapter we shall even consider finite fields $\mathbb{Z}_p$.

We shall primarily consider irreducible curves of degree 2 (*conics*) and 3 (*cubic curves*) and shall adopt the common convention of referring to the polynomial $f(x, y)$ or the equation $f = 0$ as a curve. Irreducibility is often a pain to check directly, so we will typically avoid doing so.

Examples 7.2.

1. Every degree 1 curve is irreducible: $ax + by + c = 0$ is a straight line.
2. The conic $(x + 1)^2 + y^2 - 4 = 0$ is a circle of radius 2 centered at the $(-1, 0)$.
3. The curve $(x^2 + y^2 - 1)(y - x^2) = 0$ has two irreducible components: a circle and a parabola.
4. It is conventional to insist that a curve contain *at least two* points. Consider, for example

$$\mathcal{C}_1 : x^2 + y^2 + 1 = 0 \qquad \mathcal{C}_2 : (x - 3)^2 + y^2 = 0$$

Both 'curves' are irreducible over $\mathbb{R}$, but $\mathcal{C}_1$ is empty and $\mathcal{C}_2$ contains only one point $(3, 0)$, so we don't consider these conics over $\mathbb{R}$. Over $\mathbb{C}$ however, things are more interesting:

- $\mathcal{C}_1$ is a conic: it is irreducible and contains at least two points $(\pm i, 0)$.
- $\mathcal{C}_2$ is not a conic, since it factorizes: $(x - 3)^2 + y^2 = (x - 3 + iy)(x - 3 - iy)$.

It might feel strange to use the word *curve* in $\mathbb{C}^2$, given that the *real* dimension of such an object is *two*. For instance, if we write $x = p + iq$ and $y = r + is$, and think about $\mathbb{C}^2 \cong \mathbb{R}^4$, we see that

$$x^2 + y^2 + 1 = 0 \iff p^2 - q^2 + r^2 - s^2 + 1 = 0 = pq + rs$$

$\mathcal{C}_1$ is therefore the intersection of two hypersurfaces in $\mathbb{R}^4$!

Since you've likely seen the following result in a linear algebra course, we omit the proof.

Theorem 7.3. *A real conic $ax^2 + bxy + cy^2 + dx + ey + g = 0$ may be classified by the sign of its* discriminant $\Delta := b^2 - 4ac$. *More precisely, a change of variables amounting to rotation and translation puts into one of the familiar canonical forms:*

$\operatorname{sgn}\Delta$	$+$	0	$-$
conic	*Hyperbola*	*Parabola*	*Ellipse*
canonical form	$\frac{x^2}{p^2} - \frac{y^2}{q^2} = 1$	$y^2 = 4kx$	$\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1$

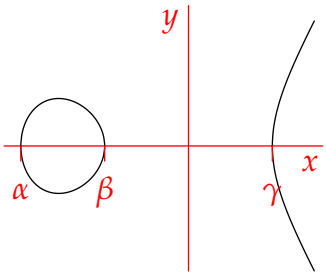
Example 7.4. The curve $f(x, y) = 7x^2 - 3xy - 2y^2 + 2x + 2$ is a hyperbola: it is irreducible, has discriminant $\Delta = 9 + 14 > 0$, and contains at least two points $(x, y) = (0, \pm 1)$.

Canonical form Cubic Curves

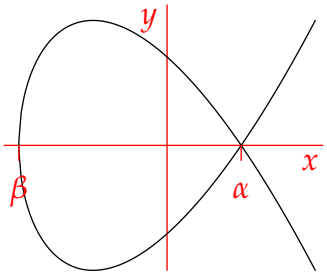
We will mostly consider cubic curves since elliptic curves will prove to be a special case. It can be shown that all (irreducible) cubic curves over $\mathbb{R}$ can be transformed to one of the four canonical forms shown. All have equations

$$y^2 = g(x) \quad \text{or equivalently} \quad f(x,y) = y^2 - g(x) = 0$$

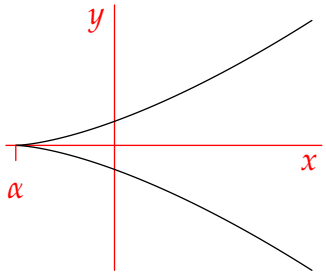
where g is a cubic polynomial in x. The classification depends on the number of distinct real roots of g.



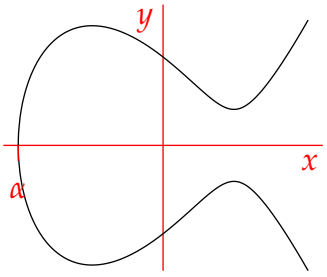
I: $y^2 = (x-\alpha)(x-\beta)(x-\gamma)$



II: $y^2 = (x-\alpha)^2(x-\beta)$



III: $y^2 = (x-\alpha)^3$



IV: $y^2 = (x-\alpha)q(x)$

In type IV, $q(x)$ is an irreducible quadratic. If we work over $\mathbb{C}$, then type IV does not exist since all quadratics may be factorized.

The classification requires more subtle changes of co-ordinates than merely rotation and translation. The details are not relevant to us since almost all examples will already be in canonical form.

Example 7.5. Fermat's curve $y^2 = x^3 - 2$ is in canonical form IV over $\mathbb{R}$ and type I over $\mathbb{C}$. If we let $\alpha = \sqrt[3]{2} \in \mathbb{R}$ and $\zeta = e^{2\pi i/3}$, then

$$y^2 = (x-\alpha)(x^2+\alpha x+\alpha^2) = (x-\alpha)(x-\alpha\zeta)(x-\alpha\zeta^2)$$

Singular Points and Non-singular Curves

In forms II and III, $g(x)$ has a repeated root at α and we call $(\alpha, 0)$ a *singular point*. More specifically:

- Type II has a *double point* where the curve crosses itself;
- Type III has a *cusp*, or *triple point*.

A better way of thinking about singular points is that the *gradient* of the curve is not well-defined.

Definition 7.6. A point p on an algebraic curve $f(x,y) = 0$ is *singular* if $\nabla f(p) = \mathbf{0}$. A curve is *non-singular* if it has no singular points.

Lemma 7.7. *1. According to the definition, only canonical forms I and IV are non-singular.*

2. At a non-singular point $p = (x_0, y_0)$ there is a well-defined (and unique) tangent line

$$f_x(p)(x-x_0) + f_y(p)(y-y_0) = 0$$

Examples 7.8. 1. Let $\mathcal{C}$ be defined by $f(x,y) = y^2 - x^3 + 7x + 6$. Then

$$\nabla f = \begin{pmatrix} -3x^2 + 7 \\ 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff x = \pm\sqrt{\frac{7}{3}},\ y = 0$$

The points $(\pm\sqrt{\frac{7}{3}}, 0)$ do not lie on the curve, so $\mathcal{C}$ is non-singular. This also follows from the fact that $\mathcal{C}$ is in canonical form I:

$$y^2 = (x+1)(x+2)(x-3)$$

2. Define $\mathcal{C}$ by $f(x,y) = x^2 + xy^2 + 2x + 1$. Then

$$\nabla f = \begin{pmatrix} 2x + y^2 + 2 \\ 2xy \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff (x,y) = (-1,0)$$

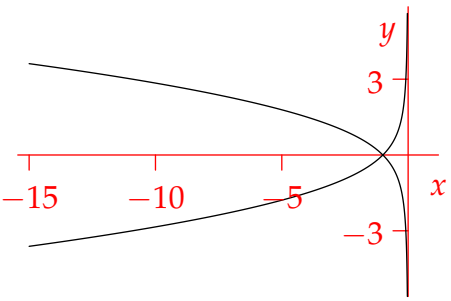
Since $f(-1,0) = 0$ it follows that $\mathcal{C}$ has one singular point. Deciding whether this is a double point or a cusp is a trickier. Consider points on a tiny circle of radius ε centered at $(-1,0)$:

$$(x,y) = (-1 + \varepsilon\cos\theta, \varepsilon\sin\theta)$$

Substituting into $f(x,y) = 0$, we obtain

$$\varepsilon^2\cos 2\theta + \varepsilon^3\cos\theta\sin^2\theta = 0$$

Since $\varepsilon^3 \ll \varepsilon^2$ we see that $\cos 2\theta = 0$, resulting in *four* nearby points corresponding to $\theta = \pm\frac{\pi}{4}, \pm\frac{3\pi}{4}$. We therefore have a double point.



Observe that $\mathcal{C}$ looks like canonical form II but with a twist: β lies at infinity. We'll pursue this in the next section.

Exercises 1. Decide whether the following curves are non-singular and compute the tangent line at $(1,2)$. For part (c), identify the type of conic.

(a) $xy^2 + 2x - 6 = 0$

(b) $2x^2 + xy - y^2 + 2x - y = 0$

(c) $x^2 - 3xy + 4y^2 + x - 5y - 2 = 0$

2. Suppose that $f(x,y) = g(x,y)h(x,y)$ and $g(p) = 0 = h(p)$. Show that p is a singular point of f.

3. Let $f(x,y) = y^2 - g(x)$. By computing ∇f, show that any singular point must be $(\alpha, 0)$ where α is a *repeated* root of g. Hence prove part 1 of Lemma 7.7: a canonical form cubic is non-singular if and only if g has distinct roots.

4. Suppose a degree two curve $\mathcal{C}$ has a singular point $p = (x_0, y_0)$.

(a) With respect to the new co-ordinates $X = x - x_0$, $Y = y - y_0$, explain why $\mathcal{C}$ has equation $aX^2 + bXY + cY^2 = 0$.

(b) Hence show that $\mathcal{C}$ is a pair of intersecting straight lines.

(c) How does the value of the discriminant Δ depend on these lines?

7.2 Homogenization and Projective Geometry

Example 7.8.2 shows that viewing cubic curves in canonical form requires us to think about points at infinity. This is even useful for quadratic curves: putting together Exercises 7.1.2 and 4 we see that all degree two curves fall in to one of the following categories:

- Conics: non-singular and irreducible.
- A pair of intersecting straight lines: singular and non-irreducible (factorizable).
- A pair of parallel lines: non-irreducible. While such have no singular points in $\mathbb{R}^2$, the idea that parallel lines meet at infinity suggests that we should also describe these curves as singular.

Definition 7.9. The *homogenization* of an algebraic curve is the equation obtained by multiplying terms by powers of a new variable until all terms have the same degree as the original curve. A homogenized curve is also called a *projective curve.*

Examples 7.10.
1. Homogenizing the circle $x^2 + y^2 = 3$ results in a projective curve $X^2 + Y^2 = 3Z^2$.
2. The homogenization of the hyperbola $x^2 - y^2 = 3$ is the projective curve $X^2 - Y^2 = 3Z^2$.
3. The homogenization of $x^2 + xy^2 + 2x + 1 = 0$ is $X^2Z + XY^2 + 2XZ^2 + Z^3 = 0$.

Definition 7.11. The *real projective plane* $\mathbb{RP}^2$ is defined via an equivalence relation on the set of non-zero points in $\mathbb{R}^3$:

$$\mathbb{RP}^2 = \mathbb{R}^3_* \Big/ \sim \quad \text{where} \quad (X,Y,Z) \sim (A,B,C) \iff \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \parallel \begin{pmatrix} A \\ B \\ C \end{pmatrix}$$

Points in $\mathbb{RP}^2$ correspond to *lines* through the origin in $\mathbb{R}^3$, and are denoted using *homogeneous coordinates* $[X,Y,Z] \in \mathbb{RP}^2$.
The set of points with $Z = 0$ is the *ideal line,* any point of which is an *ideal point.*
We may also consider the *complex projective plane* $\mathbb{CP}^2$, though typically we'll avoid this.

One can visualize $\mathbb{RP}^2$ as $\mathbb{R}^2$ together with all the 'points at infinity' on the ideal line. In particular:

Finite Points $(x,y) \in \mathbb{R}^2$ corresponds to $[x,y,1] \in \mathbb{RP}^2$. Similarly $[X,Y,Z] \rightsquigarrow (\frac{X}{Z}, \frac{Y}{Z})$ if $Z \neq 0$.

Ideal Points $[X,Y,0]$ may be visualized as a point at infinity in the direction $\pm \begin{pmatrix} X \\ Y \end{pmatrix}$.

Scaling $[X,Y,Z] = [\lambda X, \lambda Y, \lambda Z]$ for any non-zero λ.

Warning! $[0,0,0]$ does not exist and is not a valid point in $\mathbb{RP}^2$!

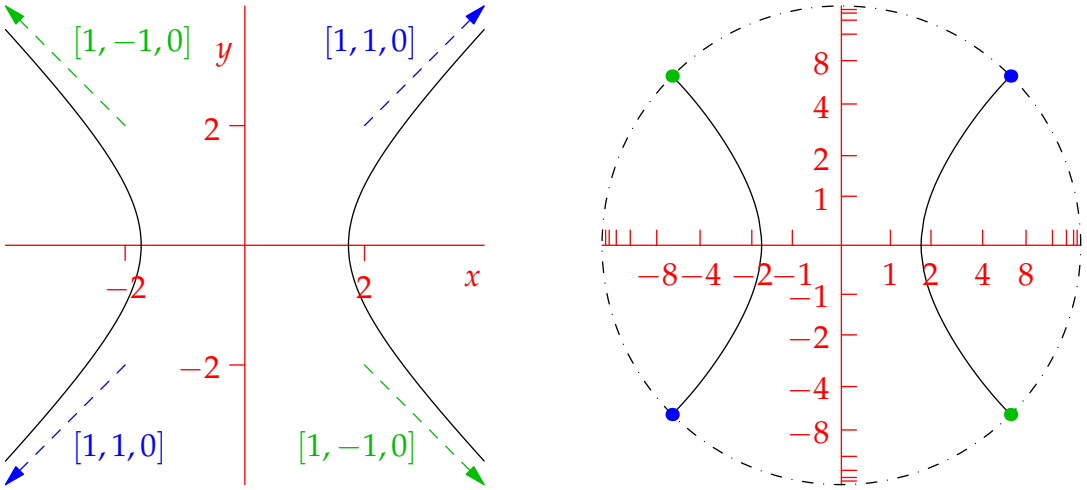
We revisit our above examples in this context.

Examples (7.10, mk. II).
1. The projective curve $X^2 + Y^2 = 3Z^2$ has no ideal points, since

$$Z = 0 \implies X^2 + Y^2 = 0 \implies X = Y = 0$$

does not produce a legitimate point in $\mathbb{RP}^2$. In $\mathbb{CP}^2$ however, the curve contain two ideal points $[1, \pm i, 0]$.

2. Given the projectivized hyperbola $X^2 - Y^2 = 3Z^2$ we see that $Z = 0 \implies X = \pm Y$ which results in two ideal points $[1, \pm 1, 0]$. This reflects the fact that the hyperbola has asymptotes $y = \pm x$.

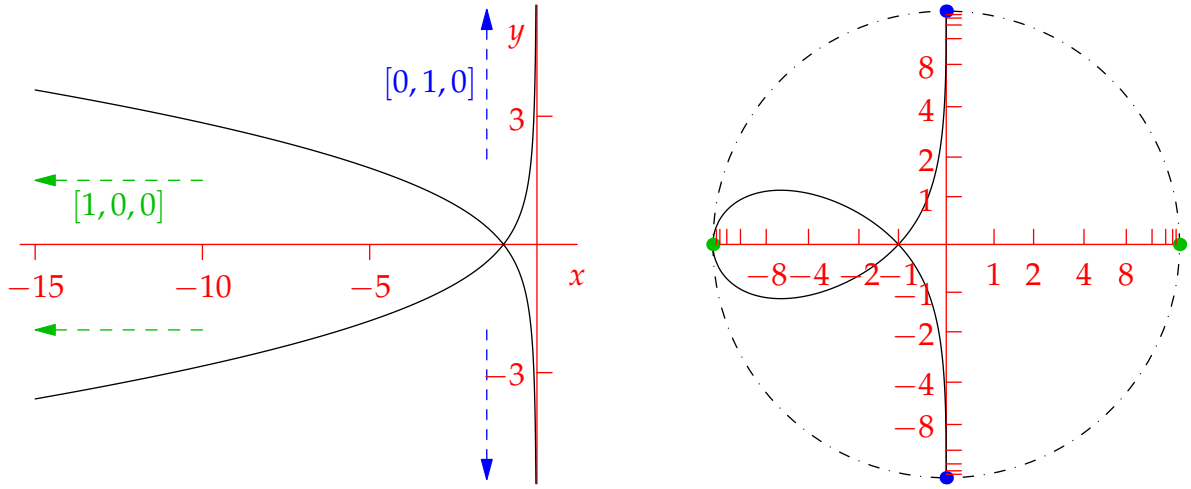


The pictures show the hyperbola in $\mathbb{R}^2$ and a representation in $\mathbb{RP}^2$ where the ideal line (dashed) can be visualised:[a] Note that $[1, 1, 0] = [-1, -1, 0]$: the blue dots represent the *same* ideal point!

3. The projective curve $X^2Z + XY^2 + 2XZ^2 + Z^3 = 0$ has two ideal points

$$Z = 0 \implies XY^2 = 0 \rightsquigarrow [1, 0, 0],\ [0, 1, 0]$$

These suggest points at infinity in the horizontal and vertical directions.



Note that the curve appears to be *tangent* to the ideal line at $[1, 0, 0]$. Our next goal is to be able to check this algebraically…

[a] Explicitly, we used $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \left(\frac{2}{\pi} \tan^{-1} \frac{\sqrt{x^2+y^2}}{3} \right) \begin{pmatrix} x \\ y \end{pmatrix}$ to map $\mathbb{R}^2$ bijectively to the inside of the unit circle.

Non-singular Projective Curves

To discuss whether ideal points are singular, we extend Definition 7.6 to the projective plane.

Definition 7.12. A projective curve $\widetilde{\mathcal{C}}$ defined by the homogeneous equation $F(X,Y,Z) = 0$ is *non-singular* at $[P]$ if and only if $\nabla F(P) \neq \mathbf{0}$. In such a case, its *tangent line* at $[P]$ has equation

$$\nabla F(P) \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = 0 \qquad \text{or} \qquad F_X(P)X + F_Y(P)Y + F_Z(P)Z = 0$$

This definition extends our previous notions of non-singularity and tangency.

Theorem 7.13. *Let $\mathcal{C}$ be an algebraic curve with homogenization $\widetilde{\mathcal{C}}$. Let $p \in \mathcal{C}$ correspond to the non-ideal point $[P] \in \widetilde{\mathcal{C}}$. Then:*

1. *$[P]$ is non-singular if and only if p is non-singular.*
2. *The homogenization of the tangent line at p is the tangent line at $[P]$.*

Instead of proving the theorem (have a go if you want to test yourself!), we work our examples using both definitions and compare. To summarize; a projective curve $\widetilde{\mathcal{C}}$ is non-singular if and only if $\mathcal{C}$ is non-singular *and* $\widetilde{\mathcal{C}}$ is non-singular at all ideal points.

Examples (7.10, mk. III). 2. The hyperbola $\mathcal{C} : f(x,y) = x^2 - y^2 - 3$ and its homogenization $\widetilde{\mathcal{C}} : F(X,Y,Z) = X^2 - Y^2 - 3Z^2$ have gradients

$$\nabla f = \begin{pmatrix} 2x \\ -2y \end{pmatrix} \qquad \nabla F = \begin{pmatrix} 2X \\ -2Y \\ -6Z \end{pmatrix}$$

The first is zero only at the origin $(x,y) = (0,0)$: this isn't on the curve, so the finite curve $\mathcal{C}$ is non-singular. Similarly, $\nabla F = \mathbf{0} \iff X = Y = Z = 0$: but $[0,0,0]$ is not a valid point, so the homogenized curve $\widetilde{\mathcal{C}}$ is also non-singular.

Now we compute the tangent lines at the corresponding points $p = (\frac{7}{4}, -\frac{1}{4})$ and $[P] = [7,-1,4]$:

$$\nabla f(p) = \begin{pmatrix} \frac{7}{2} \\ \frac{1}{2} \end{pmatrix} \implies \tfrac{7}{2}(x - \tfrac{7}{4}) + \tfrac{1}{2}(y + \tfrac{1}{4}) = 0 \implies 7x + y = 12$$
$$\nabla F(P) = \begin{pmatrix} 14 \\ 2 \\ -24 \end{pmatrix} \implies 14X + 2Y - 24Z = 0 \implies 7X + Y = 12Z$$

The tangent line in $\mathbb{RP}^2$ is plainly the homogenization of that in $\mathbb{R}^2$.

We can even compute the tangent lines at the ideal points $[1, \pm 1, 0]$: e.g.

$$\nabla F(1,1,0) = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix} \implies 2X - 2Y = 0 \implies Y = X$$

Similarly, the curve is tangent to $Y = -X$ at $[1,-1,0]$. This meshes with the fact that the hyperbola has asymptotes $y = \pm x$.

3. The curves $\mathcal{C} : f(x,y) = x^2 + xy^2 + 2x + 1$ and $\widetilde{\mathcal{C}} : F(X,Y,Z) = X^2Z + XY^2 + 2XZ^2 + Z^3$ have

$$\nabla f = \begin{pmatrix} 2x + y^2 + 2 \\ 2xy \end{pmatrix} \qquad \nabla F = \begin{pmatrix} 2XZ + Y^2 + 2Z^2 \\ 2XY \\ X^2 + 4XZ + 3Z^2 \end{pmatrix}$$

In $\mathbb{R}^2$, the first is zero only at $(x,y) = (-1,0)$: this is a singular point on the curve: recall the graph where it is clear that this is a *double point*. For the homogenization, the result is the same

$$\nabla F = \mathbf{0} \iff [X,Y,Z] = [-1,0,1]$$

We now compute the tangent lines at the ideal points:

- $\nabla F(1,0,0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \implies Z = 0$. The curve is tangent to the ideal line at $[1,0,0]$. This fits with the curve being canonical form III in disguise, with β moved to the ideal line.
- $\nabla F(0,1,0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \implies X = 0$. Again this makes sense, since the finite curve $\mathcal{C}$ is asymptotic to the line $x = 0$.

Exercises 1. Verify that the circle $x^2 + y^2 - 3 = 0$ in Example 7.10 part 1 has non-singular homogenization $X^2 + Y^2 - 3Z^2 = 0$.

2. Consider Fermat's curve $\mathcal{C} : f(x,y) = y^2 - x^3 + 2$.
   (a) Find the homeogenization $\widetilde{\mathcal{C}}$ and show that it has exactly one ideal point.
   (b) Find the equation of the tangent line at the point $(3,5)$ in two ways:
      i. By working directly with f in $\mathbb{R}^2$.
      ii. By working with the homogenized polynomial F in $\mathbb{RP}^2$.
   (c) Check that Fermat's curve is everywhere non-singular, including at its ideal point.
   (d) Show that Fermat's curve is tangent to the ideal line at its ideal point.

3. Show that the homogenization of the real quartic curve $\mathcal{C} : xy^3 + x^4 - 2 = 0$ is non-singular. Also show that it has a single ideal point and find its tangent line there.

4. Explain why proving Fermat's Last Theorem amounts to showing that, when $n \geq 3$, the only rational points on $x^n + y^n = 1$ are $(\pm 1, 0)$ and $(0, \pm 1)$.

5. (a) If a quadratic curve $\mathcal{C}$ consists of two parallel lines, show that its homogenization $\widetilde{\mathcal{C}}$ is singular at its ideal point.
   (b) Consider the projective quadratic $\widetilde{\mathcal{C}} : aX^2 + bXY + cY^2 + dXZ + eYZ + gZ^2$. Prove that

$$\widetilde{\mathcal{C}} \text{ is non-singular} \iff \det \begin{pmatrix} 2a & b & d \\ b & 2c & e \\ d & e & 2g \end{pmatrix} \neq 0$$

Together with our earlier remarks, this says that a conic is better defined as a non-singular projective curve of degree 2, and that the above determinant would make a more useful discriminant. Since the matrix is symmetric, the spectral theorem in $\mathbb{R}^3$ says it is orthogonally diagonalizable. Every projective quadratic therefore has the form $pX^2 + qY^2 + rZ^2 = 0$ in some co-ordinates with 'discriminant' pqr. Plainly p, q, r cannot have the same sign: the upshot is that all conics are equivalent in $\mathbb{RP}^2$.

7.3 Elliptic Curves & Addition

We finally(!) define our objects of study.

Definition 7.14. An *elliptic curve* over $\mathbb{Q}$ is a non-singular projective cubic with rational coefficients, containing at least one rational point.

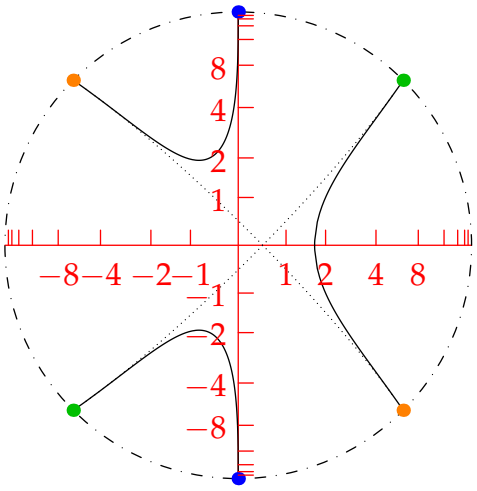
Even though an elliptic curve is *projective*, it is common to state it in standard/finite form. From now on we'll use $\mathcal{C}$ to denote curves, whether written in $\mathbb{R}^2$ or $\mathbb{RP}^2$.

Example 7.15. The curve $xy^2 + x^2 - x^3 + 2 = 0$ is elliptic. This certainly has rational coefficients and contains a rational point $(2, 1)$. To check non-singularity, compute the gradient of its homogenization $F(X, Y, Z) = XY^2 + X^2Z - X^3 + 2Z^3$:

$$\nabla F = \begin{pmatrix} Y^2 + 2XZ - 3X^2 \\ 2XY \\ X^2 + 6Z^2 \end{pmatrix}$$

is non-zero unless $X = Y = Z = 0$, which does not yield a legitimate point.

A plot in $\mathbb{RP}^2$ is shown, as are the ideal points and the tangent lines at these points:



Point	Tangent line	Dehomogenized
$[0,1,0]$	$X = 0$	$x = 0$
$[1,1,0]$	$-2X + 2Y + Z = 0$	$y = x - \frac{1}{2}$
$[1,-1,0]$	$-2X - 2Y + Z = 0$	$y = \frac{1}{2} - x$

Canonical/Weierstraß forms and the Discriminant

We will normally work with elliptic curves in canonical form; more specifically the *Weierstraß form*

$$y^2 = x^3 + cx + d$$

A simple change of co-ordinates can put any canonical form elliptic curve in Weierstraß form. It is moreover easy to check whether curves in canonical form are elliptic:

Theorem 7.16. *Suppose $\mathcal{C}$ is a canonical form cubic $y^2 = g(x)$. Then:*

1. *$\mathcal{C}$ has a unique ideal point $\sigma = [0,1,0]$, at which the ideal line $Z = 0$ is tangent. This can be taken as the rational point satisfying Definition 7.14.*
2. *Supposing g has rational coefficients, $\mathcal{C}$ is elliptic if and only if $g(x) = 0$ has distinct roots.*

We mostly leave this as an exercise except to observe that canonical forms I and IV are elliptic, and that II and III are not. To further simplify matters, we introduce a useful quantity which detects when a cubic has disinct roots.

Definition 7.17. Suppose the zeros of the cubic $g(x) = ax^3 + bx^2 + cx + d$ are $\mu_1, \mu_2, \mu_3 \in \mathbb{C}$. The *discriminant* of g is defined to be

$$\Delta := a^4 \prod_{j<k} (\mu_j - \mu_k)^2 = a^4(\mu_1 - \mu_2)^2(\mu_1 - \mu_3)^2(\mu_2 - \mu_3)^2$$

Theorem 7.18. *1. $\Delta = 0 \iff g$ has a repeated root.*[a] *In particular, a cubic curve $y^2 = g(x)$ with rational coefficients is elliptic if and only if $\Delta \neq 0$.*

2. $\Delta = (b^2 - 4ac)(c^2 - 4bd) + 2abcd - 27a^2d^2$. For reference, the common simplifications are

Monic case: $x^3 + bx^2 + cx + d \implies \Delta = (b^2 - 4c)(c^2 - 4bd) + 2bcd - 27d^2$
Weierstraß form: $x^3 + cx + d \implies \Delta = -4c^3 - 27d^2$

3. If g has integer coefficients then Δ is an integer.

[a] If the coefficients of g are real, it can also be seen that $\Delta > 0 \iff$ the cubic has distinct *real* roots.

The first is obvious from the definition. The second is extremely tedious to prove, after which the third is obvious. Only the Weierstraß form expression is worth memorizing.

Examples 7.19. 1. The curve $y^2 = x^3 - 2x + 1$ is elliptic since $\Delta = -4 \cdot (-2)^3 - 27 = 5$ is non-zero. Since $\Delta > 0$, this has canonical form I.

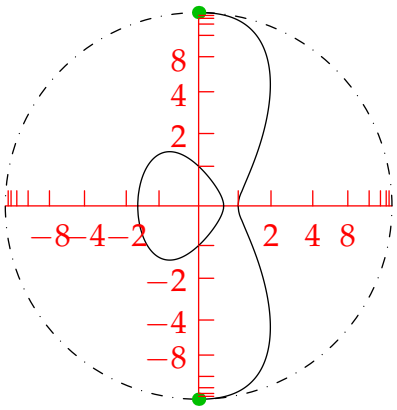
2. The curve $y^2 = x^3 - \frac{3}{4}x + \frac{1}{4}$ is not elliptic since $\Delta = -4 \cdot (-\frac{3}{4})^3 - 27 \cdot (-\frac{1}{4})^2 = 0$. Indeed this can be factored as $y^2 = (x - \frac{1}{2})^2(x + 1)$ with the repeated root clear. This has canonical form II.

3. The elliptic curve $y^2 = (x + 7)(x^2 + 1) = (x + 7)(x + i)(x - i)$ has

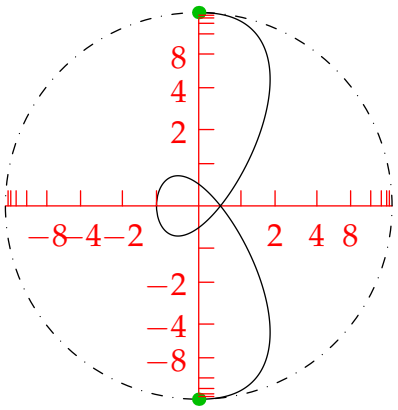
$$\Delta = (-7 - i)^2(-7 + i)^2(i + i)^2 = -10000 < 0$$

where, for variety, we used the definition of Δ.

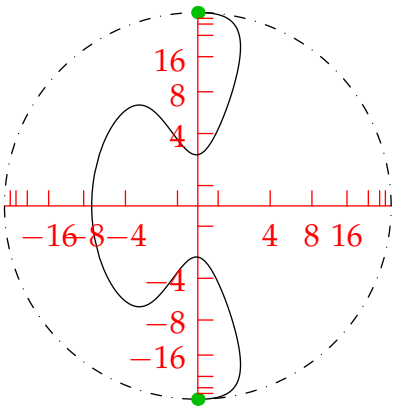
Renderings in projective space are below. The tangency of the ideal line at $\sigma = [0, 1, 0]$ should be clear. The fact that $\Delta \in \mathbb{Z}$ whenever the coefficients of g are integers will be of use to us later.



Example 1: $\Delta > 0$



Example 2: $\Delta = 0$



Example 3: $\Delta < 0$

Addition on Elliptic Curves

We now consider the property that makes elliptic curves truly special. Given p, q on a general elliptic curve $\mathcal{C}$ with marked rational point σ, here is the construction of $p+q$.

1. Construct the line ℓ joining p and q. If $q = p$, let ℓ be the tangent line at p: this exists by non-singularity.
2. Including multipicities, the set of intersections $\ell \cap \mathcal{C}$ contains exactly three points:[4] p, q and a third point $p * q$.
3. Repeat the exercise with $p * q$ and σ to define

$$p + q := (p * q) * \sigma$$

If $p = q$ we write $2p = p + p$, etc.

This construction is particularly nice if $\mathcal{C}$ is in canonical form and we choose $\sigma = [0,1,0]$. If $p * q \neq \sigma$, then the line joining it to σ is *vertical* and, by symmetry, $p + q$ is the vertical reflection of $p * q$. For instance:

Example 7.20. Let $\widetilde{\mathcal{C}}$ be the elliptic curve defined by

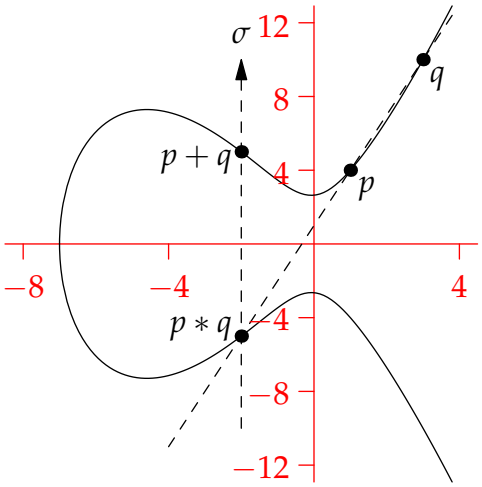
$$y^2 = (x+7)(x^2+1)$$

This contains the points $p = (1,4)$ and $q = (3,10)$. The line joining these has equation

$$y - 4 = \frac{10-4}{3-1}(x-1) \implies y = 3x + 1$$

Substituting this into the equation for the curve, we obtain

$$\begin{aligned} (3x+1)^2 &= x^3 + 7x^2 + x + 7 \\ \implies x^3 - 2x^2 - 5x + 6 &= 0 \qquad (*) \\ \implies (x-1)(x-3)(x+2) &= 0 \\ \implies p * q &= (-2,-5) \\ \implies p + q &= (-2,5) \end{aligned}$$



Computing the x-coordinate of $p * q$ (and thus $p + q$) is easier than you think since we already know two of the roots of the cubic $(*)$. In fact factorization is completely unnecessary:

Lemma 7.21. *The sum of the roots of the cubic $ax^3 + bx^2 + cx + d = 0$ is $\mu_1 + \mu_2 + \mu_3 = -\frac{b}{a}$.*

Proof. Simply multiply out $a(x-\mu_1)(x-\mu_2)(x-\mu_3) = 0\ldots$ ∎

[4] This is guaranteed to happen in projective space by *Bézout's Theorem.*

Example (7.20, continued.). Let μ_1, μ_2, μ_3 be the x co-ordinates of p, q and $p * q$. Then

$$\mu_2 = -\mu_0 - \mu_1 - \frac{b}{a} = -1 - 3 + 2 = -2$$

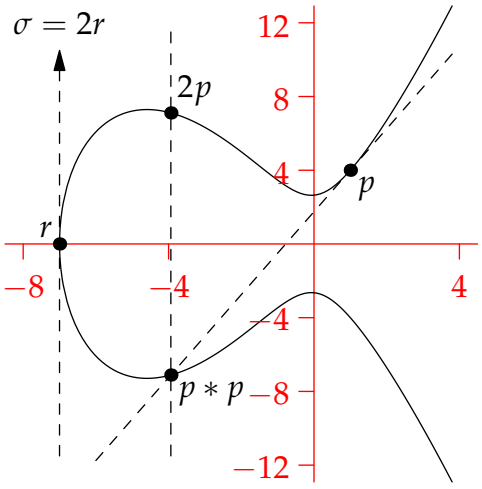
We therefore only needed to expand $(*)$ as far as the x^2 term to find the third root!
We perform another couple of additions, this time requiring tangent lines:

First compute the gradient of $f(x,y) = y^2 - x^3 - 7x^2 - x - 7$

$$\nabla f = \begin{pmatrix} -3x^2 - 14x - 1 \\ 2y \end{pmatrix}$$

At $p = (1,4)$ this yields the tangent line

$$-18(x-1) + 8(y-4) = 0 \implies y = \frac{1}{4}(9x+7)$$



Substitute into the cubic and apply the lemma:

$$\frac{81}{16}x^2 + \cdots = x^3 + 7x^2 + \cdots$$
$$\implies 2x_p + x_{p*p} = \frac{81}{16} - 7 = -\frac{31}{16} \implies x_{p*p} = -\frac{63}{16}$$

Substituting into the equation for the tangent line, we obtain $y_{p*p} = -\frac{455}{64}$. Therefore

$$2p = (p * p) * \sigma = \left(-\frac{63}{16}, \frac{455}{64}\right)$$

Finally let $r = (-7, 0)$. Since the tangent line at r is vertical, we have $r * r = \sigma$. Moreover, $\mathcal{C}$ is tangent to the ideal line $Z = 0$ at σ, whence

$$2r = (r * r) * \sigma = \sigma * \sigma = \sigma$$

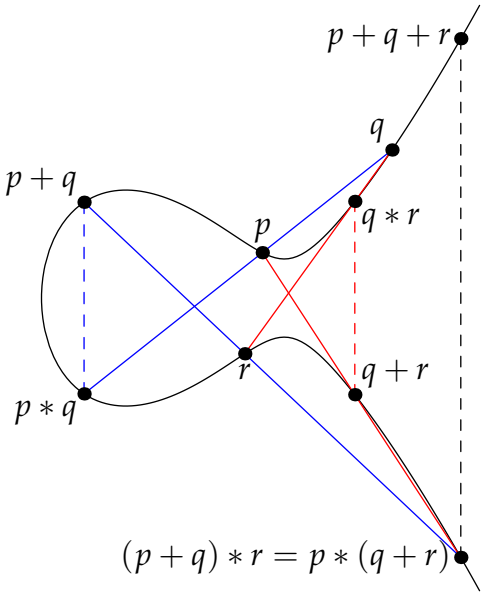
The choice of the $+$ symbol is merited by the next result.

Theorem 7.22. *If $\mathcal{C}$ is an elliptic curve, then $(\mathcal{C}, +)$ is an abelian group with identity σ.*

The identity and inverse properties are left as exercises.

Commutativity $p + q = q + p$ should be obvious since the roles of p and q are symmetric in the definition of $p * q$.

As often in group theory, the tough part is the justification of associativity $(p+q)+r = p+(q+r)$. In this case it follows fairly easily from the *Cayley–Bacharach Theorem*, a corollary of Bézout's Theorem. The details would take us too far afield, so the picture will suffice as an example.



Exercises 1. Use the discriminant to check that the curve $y^2 = (x+2)\big(8(x-1)^2+1\big)$ is elliptic.

2. Check that the curve $y^2 = x^3 - x + 1$ is elliptic. If $p = (1,1)$, verify that

$$2p = (-1,1), \quad 3p = (0,-1), \quad 4p = (3,-5), \quad 5p = (5,11), \quad 6p = (\tfrac{1}{4}, \tfrac{7}{8})$$

3. Given $p = (-1,4)$ and $q = (2,5)$ on the curve $y^2 = x^3 + 17$, show that

$$p + q = (-\tfrac{8}{9}, -\tfrac{109}{27}), \quad 2p = (\tfrac{137}{64}, -\tfrac{2651}{512})$$

4. (a) Suppose that $y^2 = ax^3 + bx^2 + cx + d$ is a cubic. Let $x = a(u - \frac{b}{3a^2})$ and $y = a^2 v$ to transform the curve to Weierstrass form (i.e. $a = 1, b = 0$).

   (b) Suppose now that $y^2 = x^3 + cx + d$ where $c, d \in \mathbb{Q}$. Let k be the least common multiple of the denominators of c and d. Show that $u = k^2 x$ and $v = k^3 y$ transform the curve to a Weierstraß form with *integer coefficients.*

5. Suppose $\mathcal{C} : y^2 = ax^3 + bx^2 + cx + d$ is an elliptic curve in canonical form with $\sigma = [0,1,0]$. Given finite points $p = (x_0, y_0)$ and $q = (x_1, y_1)$ on $\mathcal{C}$ where $x_0 \neq x_1$, use Lemma 7.21 to prove that

$$x_{p+q} = \frac{m^2 - b}{a} - x_0 - x_1 \quad \text{and} \quad y_{p+q} = m(x_0 - x_{p+q}) - y_0 \quad \text{where} \quad m = \frac{y_1 - y_0}{x_1 - x_0}$$

   Now find an expression for the co-ordinates of $2p$ whenever $y_0 \neq 0$.

   These expressions can be easily be fed to a computer for rapid computation.

6. We prove Theorem 7.16 (and a bit more). Let $g(x)$ be a polynomial of degree n.

   (a) Prove that $g(x) = 0$ has a repeated root at $x = \mu$ if and only if $g(\mu) = g'(\mu) = 0$.
   (*Hint: write $g(x) = (x-\mu)^k h(x)$, where k is the multiplicity of the root $x = \mu$*)

   (b) Let $\mathcal{C}$ be defined by $f(x,y) = y^2 - g(x)$. Compute the gradient of f and use (a) to show that all finite points on $\mathcal{C}$ are non-singular if and only if g has distinct roots.

   (c) If $n = 3$, argue that the cubic $\mathcal{C}$ contains a unique ideal point $\sigma = [0,1,0]$ and that the curve is tangent to the ideal line at σ.

7. Suppose that $\mathcal{C}$ is an elliptic curve with marked point σ. We establish the identity and inverse properties for the group operation $+$.

   (a) (Identity) The point $p + \sigma = (p * \sigma) * \sigma$ lies on the intersection of $\mathcal{C}$ and the line ℓ joining $p * \sigma$ and σ. Explain why $\ell \cap \mathcal{C} = \{p, \sigma, p * \sigma\}$ and thus why $p + \sigma = p$.
   (*Answer for canonical form curves if you like, though this is not necessary and σ need not be $[0,1,0]$. Indeed you may find it easier to draw pictures if σ is finite!*)

   (b) (Inverse) Let ℓ be the tangent line to $\mathcal{C}$ at σ and let $\tau \in \mathcal{C}$ be such that $\ell \cap \mathcal{C} = \{\sigma, \tau\}$. For any $p \in \mathcal{C}$, prove that $p + (p * \tau) = \sigma$, whence $-p = p * \tau$.
   (*If $\mathcal{C}$ is in canonical form and $\sigma = [0,1,0]$, then $\tau = \sigma$ and $-p = p * \sigma$: the tangent at σ has a triple intersection!*)

   The whole discussion can be extended to general cubic curves: the non-singular points form a group.

8. $\mathcal{C} : y^2 = x^3 - 3x + 2 = (x+2)(x-1)^2$ is not an elliptic curve. Suppose that $p, q \in \mathcal{C}$ such that both $p, q \neq (1,0)$. Prove that $p + q \neq (1,0)$.
   (*Hint: think about all lines through $(-1,0)\ldots$*)

7.4 Rational Points on Elliptic Curves

A standard problem when given an elliptic curve is to find all the rational points lying on it. There may be very few, or infinitely many. For example:

1. Fermat's curve $x^3 + y^3 = 1$ famously contains only two finite rational points $p = (1, 0)$ and $q = (0, 1)$. Together with the unique ideal point $\sigma = [1, -1, 0]$, these form a group isomorphic to $\mathbb{Z}_3$: it is nice exercise to check that $q = 2p$ and $\sigma = 3p$, etc.

2. The curve $y^2 = x^3 + 17$ can be shown to contain infinitely many rational points. Indeed the point $p = (-1, 4)$ generates an infinite sequence of distinct rational points $p, 2p, 3p, 4p, 5p, \ldots$

The examples suggest the following:

Corollary 7.23. *If p and q are rational points on an elliptic curve, so also is $p + q$. Moreover, the set of rational points on the curve $\mathcal{C}_\mathbb{Q}$ forms an abelian group under $+$.*

Proof. The line ℓ through p and q has rational coefficients: this is also true when $p = q$ and ℓ is the tangent line. Substitute ℓ into $\mathcal{C}$ to obtain a cubic with rational coefficients and two rational roots (the x or y co-ordinates of p and q). The third root, a co-ordinate of $p * q$, is rational by Lemma 7.21: the other co-ordinate of $p * q$ is rational via ℓ.

Since σ is rational, the same applies to $p + q = (p * q) * \sigma$ and, by Exercise 7.3.7, to $-p$. We conclude that $\mathcal{C}_\mathbb{Q}$ is a subgroup of $\mathcal{C}$. ∎

For a curve in canonical form, Exercise 7.3.5 shows the rationality of $p + q$ explicitly.

An early key result regarding the group $\mathcal{C}_\mathbb{Q}$ was proved in the 1920's.

Theorem 7.24 (Mordell). *$\mathcal{C}_\mathbb{Q}$ is finitely generated: all rational points can be produced from a finite subset using only the operation $+$. It follows[a] that $\mathcal{C}_\mathbb{Q} \cong T \times \mathbb{Z}^r$ for some finite subgroup T (the* torsion *subgroup) and $r \in \mathbb{N}_0$ (the* rank *of $\mathcal{C}$).*

[a] This is just the fundamental theorem of finitely generated abelian groups applied to $\mathcal{C}_\mathbb{Q}$.

The proof of Mordell's theorem is accessible, but too involved for us.[5] Computing the rank of an elliptic curve is exceedingly difficult and is the topic of much research. For instance, it is conjectured that roughly half of elliptic curves have rank zero, half have rank 1, and a relatively tiny number have rank ≥ 2.

By contrast, and as we'll see below, it is often very easy to find the torsion subgroup. Note that being in the torsion subgroup means that a point has *finite order*: $\exists m \in \mathbb{N}$ such that $mp = \sigma$.

Example 7.25. As we saw above, the curve $x^3 + y^3 = 1$ has $\mathcal{C}_\mathbb{Q} = \{\sigma, (1, 0), (0, 1)\}$. Clearly $\mathcal{C}$ has rank zero and torsion $T \cong \mathbb{Z}_3$

Plainly σ is the only element of order 1 on $\mathcal{C}_\mathbb{Q}$. Our next result shows how few possibilities there are for the finite order of an element.

[5] The 'size' of a rational point on $\mathcal{C}$ must be suitably defined, then a descent argument is performed showing how rational points may be produced from those with smaller size. Look it up if you are interested.

Theorem 7.26 (Mazur 1977/8). *Suppose a rational point on an elliptic curve $\mathcal{C}$ has finite order k. Then $1 \le k \le 10$ or $k = 12$. Moreover, the torsion of $\mathcal{C}_{\mathbb{Q}}$ is isomorphic to one of the following:*

- $\mathbb{Z}_n$ *where* $n = 1, 2, \ldots 10$ *or* 12*;*
- $\mathbb{Z}_2 \times \mathbb{Z}_{2n}$ *where* $n = 1, 2, 3$ *or* 4*.*

The next result provides a method for computing the torsion using the discriminant (Definition 7.17).

Theorem 7.27 (Nagell–Lutz 1935/7). *Suppose (x, y) is a rational point of finite order on the elliptic curve $y^2 = g(x) = x^3 + bx^2 + cx + d$, where $b, c, d \in \mathbb{Z}$. Then $x, y \in \mathbb{Z}$ and either $y = 0$ or $y^2 \mid \Delta$.*

Here is the process for finding the torsion.

- Compute Δ and find all integers $y = 0$ or satisfying $y^2 \mid \Delta$.
- If there exists an *integer* x satisfying $p = (x, y) \in \mathcal{C}$, keep the point and compute $2p$, $3p$, etc.
- If $np = (x, y)$ ever has $x, y \notin \mathbb{Z}$ or $y^2 \nmid \Delta$, stop: p does not have finite order.
- If $np = \sigma$ for some n, then p has finite order. By Mazur's Theorem, you don't have to look far.

Before giving the proof, here are several examples. We also state the rank of each curve.

Examples 7.28. 1. Consider the elliptic curve $y^2 = x^3 + 1$ with discriminant $\Delta = -27$.

Since 3 is the only prime dividing Δ, we see that a rational point $p = (x, y)$ of finite order must have $y = 0, \pm 1$ or ± 3. These all yield points on the curve:

$$p := (-1, 0), \quad \pm q := (0, \pm 1), \quad \pm r := (2, \pm 3)$$

The relationship should be obvious from the picture, but we compute to make sure. With $f(x, y) = y^2 - x^3 - 1$, find the tangent line at r:

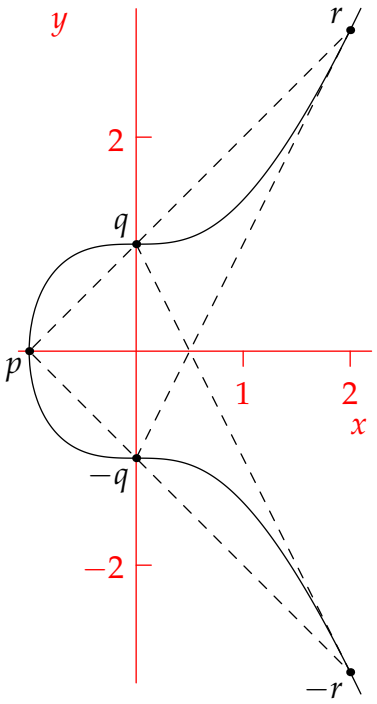
$$\nabla f(r) = \begin{pmatrix} -12 \\ 6 \end{pmatrix} \implies -2(x-2) + (y-3) = 0$$

Substituting into the cubic yields

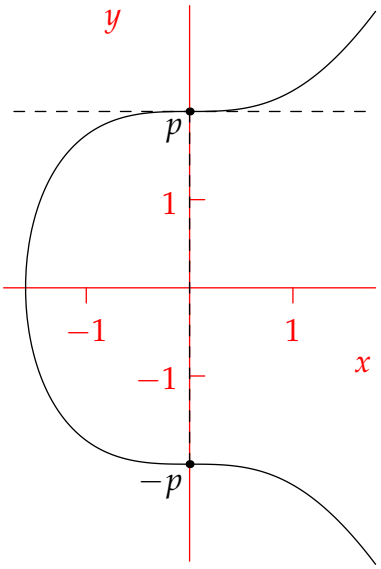
$$\begin{aligned} x^3 - 4x^2 + \cdots = 0 &\implies x_r + x_r + x_{r*r} = 4 \\ &\implies 2r = (0, 1) = q \end{aligned}$$

It can similarly be checked that p, q, r have orders 2, 3 and 6 respectively. Together with σ, the torsion is isomorphic to $\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3$.

This example has rank zero: there are no other rational points on the curve! If you want a challenge...



2. The curve $y^2 = x^3 + 4$ has discriminant $\Delta = -27 \cdot 4^2$.



For points (x, y) of finite order Nagell–Lutz says that $y = 0, \pm 1, \pm 2, \pm 3, \pm 4, \pm 6$, or ± 12. However, the only integer points with such are $\pm p := (0, \pm 2)$.

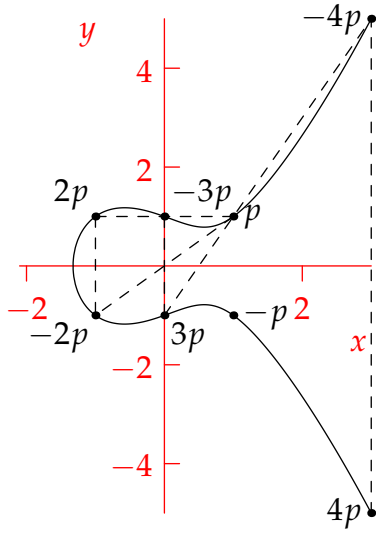
With $f(x, y) = y^2 - x^3 - 4$, find the tangent line at $p = (0, 2)$:

$$\nabla f(p) = \begin{pmatrix} 0 \\ 4 \end{pmatrix} \implies 4(y - 2) = 0$$

which is the horizontal line $y = 2$. Substituting into equation $y^2 = x^3 + 4$ yields $x^3 = 0$, whence $p * p = p$ and $2p = -p = (0, -2)$. Since p and $2p$ are vertically aligned, we have $3p = 2p + p = \sigma$. Clearly p and $-p$ both have order three: together with σ, the torsion is isomorphic to $\mathbb{Z}_3$.

The rank is again zero: the curve has only three rational points.

3. The converse to Nagell–Lutz is *false*: there could be points (x, y) on $\mathcal{C}$ with $x, y \in \mathbb{Z}$ and $y^2 \mid \Delta$, but where (x, y) has *infinite order.* Exercise 7.3.2 provides such.



If $y^2 = x^3 - x + 1$ then $\Delta = -23$. To find the rational points of finite order it suffices to check $y = \pm 1$. Taking $p = (1, 1)$ we obtain, $2p = (-1, 1), 3p = (0, -1), 4p = (3, -5)$. Since $(-5)^2 \nmid \Delta$, we see that $4p$ does not have finite order, and so neither does p.

Repeating the exercise with $(1, -1)$ simply switches the sign of all y co-ordinates. The curve is therefore *torsion-free*: $T = \{\sigma\}$ is trivial.

It is significantly harder to show that the rank of this curve is 1: indeed $p = (1, 1)$ is a generator and $\mu : \mathbb{Z} \to \mathcal{C}_{\mathbb{Q}} : m \mapsto mp$ is an isomorphism.

4. The curve $y^2 = x^3 + 17$ is torsion-free and has rank two: $p = (-2, 3)$ and $q = (-1, 4)$ are generators, whence $\mu : \mathbb{Z}^2 \to \mathcal{C}_{\mathbb{Q}} : (m, n) \mapsto mp + nq$ is an isomorphism. Check the claim about being torsion-free yourself.

5. Consider the elliptic curve $y^2 = \frac{1}{2}x^3 + \frac{1}{4}$. To apply Nagell–Lutz we first have to put this in the correct form. Multiplying by 16, we obtain

$$16y^2 = 8x^3 + 4 \implies (4y)^2 = (2x)^3 + 4 \implies v^2 = u^3 + 4$$

where $(u, v) = (2x, 4y)$. Clearly the rational points on the two curves correspond. Moreover, the transformation is linear, so straight lines map to straight lines and nothing about the group operation is altered. It follows that the curve has rank zero and that the torsion is simply the translation of that of $y^2 = x^3 + 4$, namely

$$T = \{\sigma, (0, \tfrac{1}{2}), (0, -\tfrac{1}{2})\}$$

Alternative versions of Nagell–Lutz are available for elliptic curves in other standard forms. If you want more examples, you can find thousands here with explanations of terminology[6] For instance, try typing $[0,0,0,-1,1]$ or $[0,0,0,0,17]$ into the search box...

Sketch Proof of Nagell–Lutz. 1. The difficult part is showing that any rational point of finite order is an integer point. This is somewhat involved. Once this is out of the way...

2. If $p = (x,y) \in \mathcal{C}$, then Exercise 7.3.5 shows that $2p$ has x co-ordinate

$$\frac{x^4 - 2cx^2 - 8dx + c^2 - 4bd}{4(x^3 + bx^2 + cx + d)} =: \frac{h(x)}{4g(x)} = \frac{h(x)}{4y^2}$$

3. There exist polynomials $G(x), H(x)$ with integer coefficients for which

$$\Delta = G(x)g(x) + H(x)h(x)$$

Indeed $G(x) = 3x^3 - bx^2 - 5cx + 2bc - 27d$ and $H(x) = -3x^2 - 2bx + b^2 - 4c$ are such.

4. If $y \neq 0$ and p has finite order, then $2p$ also has finite order whence both have integer co-ordinates. It follows from part 2 that

$$4y^2 \mid h(x) \implies y^2 \mid h(x)$$

Since $y^2 = g(x)$, part 3 says that $y^2 \mid \Delta$.

■

Exercises 1. Suppose that a point p on an elliptic curve has finite order n: i.e. $np = \sigma$. Prove that any multiple kp also has finite order.

2. (a) For each of the following points p, q on the elliptic curve $\mathcal{C} : y^2 = x^3 + 17$, compute $p + q$.
   i. $p = (-1,4), q = (2,-5)$.
   ii. $p = (43,282), q = (52,-375)$.
   iii. $p = (-2,3), q = (\frac{19}{25}, \frac{522}{125})$
   (b) Use Nagell–Lutz to prove that *none* of the above points have finite order.

3. (a) Find the discriminant of the curve $y^2 = x^3 - 2x + 1$.
   (b) Find all the rational points of finite order on $y^2 = x^3 - 2x + 1$.

4. Let $\mathcal{C}$ be the elliptic curve $y^2 = x^3 - 43x + 166$. The points $p = (3,8)$ and $q = (-5,6)$ form part of the torsion collection of $\mathcal{C}_{\mathbb{Q}}$. Find the full collection.
   (*Hint: this is hard if you work straight from the discriminant* $\Delta = -2^{15} \cdot 13$ *as there are lots of values to check. Instead, show that* $2p = -q$*, define* $r = -3p$ *and show that* $4p = r$*. What is the order of* p*? What does Mazur's theorem tell you about the torsion subgroup?*)

5. If $\mathcal{C}$ is an elliptic curve in canonical form, show that there are exactly three points of order two on the complexified curve $\mathcal{C}_{\mathbb{C}}$. Hence show that there are either none, one or three *rational* points of order 2.

[6] For technical reasons involving the primes 2 and 3, their canonical forms are sometimes a little different to ours, and their discriminant is ours multiplied by 16.

6. Any curve 'birationally equivalent' to an elliptic curve is also called elliptic. Given two curves $\mathcal{C}_1, \mathcal{C}_2$ with polynomials f, g, this means that there exist rational functions ϕ_i, ψ_i (with rational coefficients) for which

$$s = \phi_1(x,y), \quad t = \phi_2(x,y), \qquad x = \psi_1(s,t), \quad y = \psi_2(s,t)$$

gives a bijection between all but finitely many points on $\mathcal{C}_1, \mathcal{C}_2$.

(a) Check that Fermat's curve $x^3 + y^3 = 1$ is birationally equivalent to the Weierstrass form curve $t^2 = s^3 - 432$ via the transforms

$$s = \frac{12}{x+y}, \quad t = \frac{36(x-y)}{x+y}, \quad x = \frac{6}{s} + \frac{t}{6s}, \quad y = \frac{6}{s} - \frac{t}{6s}$$

What points, if any, on $\mathcal{C}_1, \mathcal{C}_2$ are excluded from the bijection?

(b) Given that rational points must be mapped to rational points by a birational transformation, find all the rational points on the curve $t^2 = s^3 - 432$.

(*Hint: How many rational points lie on Fermat's curve?*)

7. The discriminant of each of the (non-elliptic) curves $\mathcal{C}^1 : y^2 = x^3$ and $\mathcal{C}^2 : y^2 = x^3 + x^2$ is zero.

(a) Graph these curves. Each has one singular point: what is it?

(b) Consider lines through the singular point $p \in \mathcal{C}_1$. Show that $\mathcal{C}_1$ contains infinitely many rational points (indeed it has infinitely many *integer* points).

(c) It can be seen that $(\mathcal{C}_1 \setminus \{p\}, +, \sigma)$ is a group where $\sigma = [0,1,0]$. Define $\phi : \mathcal{C}^1_{\mathbb{Q}} \setminus \{p\} \to \mathbb{Q}$ by

$$\phi(x,y) = \frac{x}{y}, \quad \phi(\sigma) = 0$$

Prove that ϕ is a bijection and that it maps inverses in $\mathcal{C}$ to inverses in $(\mathbb{Q}, +)$.

It can be shown that ϕ is a homomorphism and thus an isomorphism of groups.[a] *Since $(\mathbb{Q}, +)$ is* not *finitely generated, Mordell's theorem does not hold for singular cubic curves. It can similarly be shown that $\mathcal{C}^2_{\mathbb{Q}} \setminus \{p\} \cong (\mathbb{Q}^\times, \cdot)$ which is not finitely generated either!*

[a] If you want a challenge, prove this: try to show that any point (x,y) on the cubic has $\phi = \frac{x}{y}$ satisfying a cubic equation with no quadratic term. What does this say about three collinear points $p, q, p * q$ on $y^2 = x^3$?

7.5 Algebraic Curves Modulo a Prime

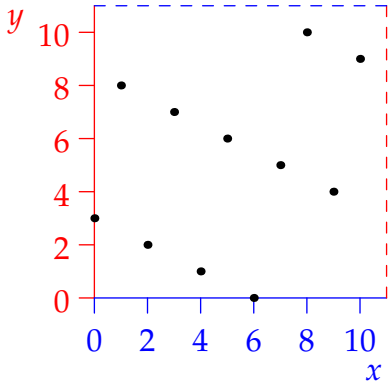
An algebraic curve with integer coefficients can be viewed as a 'curve' over the finite field $\mathbb{Z}_p$. This isn't really a curve in the traditional sense since there can only be finitely many points on $\mathcal{C}_p$. This also means that finding points relatively easy: just test all p^2 possibilities!

Definition 7.29. Given an algebraic curve $\mathcal{C} : f(x,y) = 0$ with integer coefficients, its *reduction* modulo p is the set

$$\mathcal{C}_p = \{(x,y) \in \mathbb{Z}_p^2 : f(x,y) = 0\} \qquad \text{(equivalently } f(x,y) \equiv 0 \pmod{p}\text{)}$$

The number of points on $\mathcal{C}_p$ is denoted N_p, and the *defect* is $a_p := p - N_p$.

Example 7.30. The curve $2x + 4y = 1$ over $\mathbb{Z}_{11}$ is plotted.



Observe that $N_{11} - 11$: conincidence? Since $4^{-1} = 3$ modulo 11, we can multiply by 3 to solve for y in terms of x:

$$2x + 4y = 1 \iff 6x + y = 3 \iff y = 3 - 6x = 3 + 5x$$

The same thing happens modulo *any prime except 2*:

$$x = 2^{-1}(1 - 4y) \implies N_p = p$$

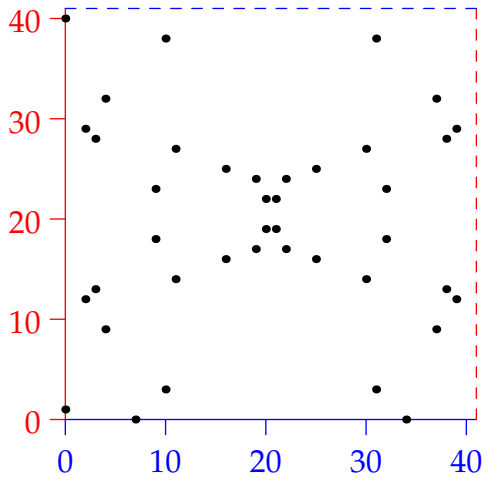
Modulo 2, there are no solutions $N_2 = 0$.

The example is easily generalized: if $b \neq 0$ in $\mathbb{Z}_p$, then $ax + by = c \iff y = b^{-1}(c - ax)$.

Quadratic Curves A simple parabola such as $y = x^2 + 7x - 3$ plainly has $N_p = p$ since y is given as a function of x. However, things are not always so clear-cut.

Consider the curve $\mathcal{C} : y^2 - 5x^2 = 1$ over $\mathbb{Z}_p$. Here are the solutions for the first few primes.

p	Solutions (x,y)	N_p	a_p
2	$(1,0), (0,1)$	2	0
3	$(0,1), (0,2), (1,0), (2,0)$	4	-1
5	$(x,1), (x,4),\ \forall x$	10	-5
7	$(\pm 2, 0), (0, \pm 1), (3, \pm 2), (4, \pm 2)$	8	-1
11		10	1
13		14	-1
17		18	-1
19		18	1
23		24	-1
29		28	1
31		30	1
37		38	-1
41		40	1
997		998	-1



Points on $y^2 - 5x^2$ modulo 41

Once we get past $p = 5$, there seems to be a pattern: think about each prime modulo 5…

To establish the pattern we count N_p similarly to how we parametrized rational points on real quadratics, by intersecting lines with $\mathcal{C}_p$.

Suppose $p \neq 2, 5$. Start with the base point $Q := (0, 1)$ which lies in $\mathcal{C}_p$ independently of p. For each $m \in \mathbb{Z}_p$, intersect $\mathcal{C}_p$ with the line $y = mx + 1$ through Q to obtain

$$0 = (m^2 - 5)x^2 + 2mx + 1 - 1 = x\left[(m^2 - 5)x + 2m\right] \qquad (*)$$

Since p is prime, there are *at most two* solutions to this equation and thus at most two intersections:

1. $x = 0$ corresponds to the base point $Q = (0, 1)$ and the 'slope' $m = 0$;
2. $m^2 \neq 5 \implies (x, y) = \left(\frac{2m}{5-m^2}, \frac{5+m^2}{5-m^2}\right) \in \mathcal{C}_p$ where $\frac{1}{5-m^2} := (5 - m^2)^{-1} \in \mathbb{Z}_p$.

Conversely, if $a \neq 0$, then the line joining Q and $(a, b) \in \mathcal{C}_p$ has slope $m = a^{-1}(b - 1) \in \mathbb{Z}_p$:

$$a(y - 1) = (b - 1)x \leftrightsquigarrow y = a^{-1}(b - 1)x + 1$$

The only point missing from this construction is $(0, p - 1)$, which corresponds to a 'vertical' line through Q.

Putting everything together, we count the number of points N_p:

- In addition to $(0, p - 1)$, each $m \in \mathbb{Z}_p$ generates a point $(x, y) \in \mathcal{C}_p$, for a base count of $p + 1$.
- If 5 is a quadratic residue modulo p, then there are *two* solutions $(\pm m)$ to $m^2 = 5 \in \mathbb{Z}_p$: for these m we cannot define $(5 - m^2)^{-1}$ and so our count rduces to $p - 1$. By quadratic reciprocity

$$\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 5 \\ -1 & \text{if } p \equiv \pm 2 \pmod 5 \end{cases}$$

In conclusion: if $p \neq 2, 5$ is prime, then the curve $y^2 - 5x^2 = 1$ modulo p has N_p points, where

$$N_p = \begin{cases} p - 1 & \text{if } p \equiv \pm 1 \pmod 5 \\ p + 1 & \text{if } p \equiv \pm 2 \pmod 5 \end{cases} \qquad \left(\text{equivalently} \quad a_p = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 5 \\ -1 & \text{if } p \equiv \pm 2 \pmod 5 \end{cases}\right)$$

Our analysis fails in $\mathbb{Z}_2$ and $\mathbb{Z}_5$. Here is why:

- If $p = 2$, then $(0, p - 1) = (0, 1)$ is not a new point. Moreover $(*)$ degenerates to $(m^2 - 1)x^2 = 0$.
- If $p = 5$, then $(*)$ becomes $mx(mx + 2) = 0$. Choosing $m = 0$ results in *five* points!

A little more is going on: modulo these primes, the 'curves' become *singular*:

$$\nabla(Y^2 - 5X^2 - Z^2) = \begin{pmatrix} -10X \\ 2Y \\ -2Z \end{pmatrix} = \mathbf{0} \iff 10X = 0 = 2Y = -2Z$$

Modulo 2, all points are singular. Modulo 5, the ideal point $[1, 0, 0]$ is singular. Relatedly, both curves factorize:

$$y^2 - 5x^2 \equiv 1 \pmod 2 \implies (x + y + 1)^2 \equiv 0 \pmod 2$$
$$y^2 - 5x^2 \equiv 1 \pmod 5 \implies (y + 1)(y - 1) \equiv 0 \pmod 5$$

The primes $p = 2, 5$ are known as *primes of bad reduction*, or simply *bad primes*. It should be clear that modulo every other prime, the curve $\mathcal{C}_p$ is non-singular.

Elliptic curves and reductions modulo p

We could continue to play with quadratic curves, but the above example is somewhat generic. Elliptic curves are more interesting. To keep things simple, we'll work only with elliptic curves in monic canonical form $y^2 = g(x) = x^3 + bx^2 + cx + d$ where $b, c, d \in \mathbb{Z}$.

Definition 7.31. A *prime of bad reduction* is a prime p such that $y^2 = g(x)$ is singular modulo p. We take this to mean that the curve contains at least one point $Q = (x_0, y_0) \in \mathbb{Z}_p^2$ for which

$$\nabla f(Q) = \begin{pmatrix} -g'(x_0) \\ 2y_0 \end{pmatrix} = \mathbf{0} \qquad (\dagger)$$

Alternatively, p is a *good prime* if $\mathcal{C}$ remains an elliptic curve over the finite field $\mathbb{Z}_p$. The construction of $P + Q$ survives the reduction: by Lemma 7.21, if P, Q are integer points, so also is $P + Q$. It follows that $\mathcal{C}_p$, when combined with $\sigma = [0, 1, 0]$, is a finite abelian group. Here is the key result.

Theorem 7.32. *p is a bad prime if and only if $p = 2$ or $p \mid \Delta$.*

Proof (sketch). Suppose $y^2 = x^3 + cx + d$ is in Weierstraß form and let $Q = (x_0, y_0) \in \mathcal{C}_p$. If $p = 2$, then (†) says that Q is singular if and only if

$$g'(x_0) = -3x_0^2 - c = x_0^2 - c = 0 \in \mathbb{Z}_2$$

$Q = (c, g(c))$ is easily checked to be a singular point on the curve.

Now assume $p \geq 3$. The rest of the proof follows from two easily verifiable polynomial identities:

1. $\Delta = -4c^3 - 27d^2 = (18cx - 27d)g(x) + (-18cx^2 + 9dx - 4c^2)g'(x)$
2. $4c^3 g(x) = (2cx + 3d)^2(cx - 3d) - \Delta(cx + d)$

If Q is singular, then $\nabla f(Q) = \mathbf{0} \implies p \mid g'(x_0)$ and $p \mid 2y_0$ from which $p \mid y_0$. But then $p \mid g(x_0)$ and identity 1 tells us that $p \mid \Delta$.

Now suppose $p \mid \Delta$. Identity 2 says that $4c^3 g(x) = (2cx + 3d)^2(cx - 3d)$ in $\mathbb{Z}_p$. There are two cases:

- If $p \nmid c$, then $g(x) = (4c^3)^{-1}(2cx + 3d)^2(cx - 3d) \implies g(x) = 0$ has a repeated root $x_0 = -(2c)^{-1}(3d) \in \mathbb{Z}_p$. But then (Exercise 7.3.6) $g'(x_0) = 0$ and we conclude that $(x_0, 0)$ is singular.
- If $p \mid c$, then $g(x) = x^3 + d$ and $\Delta = -27d^3$ modulo p, from which $p = 3$ or $p \mid d$. Note that

  $$\nabla f = \begin{pmatrix} -3x^2 \\ 2y \end{pmatrix} \in \mathbb{Z}_p^2$$

  If $p = 3$, then $x^3 = x$ for all x and so $(-d, 0)$ is a singular point on the curve. If $p > 3$ and $p \mid d$, then $(0, 0)$ is singular.

■

In a proof for non-Weierstraß curves, the main difference is that the polynomial identities are uglier.

Examples 7.33. 1. The curve $y^2 = x^3 + 1 = g(x)$ has $\Delta = -27$ with bad primes $p = 2$ and 3. Computing $\nabla(y^2 - x^3 - 1) = \begin{pmatrix} -3x^2 \\ 2y \end{pmatrix}$, we see that following points are singular:

- $p = 2$: $(x, y) = (c, g(c)) = (0, 1)$;
- $p = 3$: $(x, y) = (-d, 0) = (2, 0)$. Since $y^2 = x^3 + 1 = (x+1)(x^2 - x + 1) = (x+1)^3$ over $\mathbb{Z}_3$, the singular curve has a cusp.

2. $y^2 = x^3 - x - 6 = (x-2)(x^2 + 2x + 3)$ has $\Delta = -4(-1)^3 - 27(-6)^2 = -968 = -2^3 \cdot 11$. Computing $\nabla(y^2 - x^3 - 1) = \begin{pmatrix} -3x^2+1 \\ 2y \end{pmatrix}$, we find the following singular points:

- $p = 2$: $(x, y) = (c, g(c)) = (1, -6) = (1, 0)$;
- $p = 11$: Over $\mathbb{Z}_{11}$ the curve has a double point $y^2 = (x-2)^2(x+4)$: this corresponds to the proof in that we have a repeated root for $g(x) = 0$ at

$$x_0 = -(2c)^{-1}(3d) = -(-2)^{-1}(-18) = 2^{-1} \cdot 4 = 2 \in \mathbb{Z}_{11}$$

Elliptic Curve Cryptography (non-examinable)

Before considering the number of points on a curve modulo p, here is an easy to follow application.

Alice wishes to send a secret message to Bob. They start by agreeing on a point P on an elliptic curve $\mathcal{C}$ with 'good' prime q. In practice this is public information: software implementing the algorithm will often select a suitable curve from a database such as that maintained by NIST (e.g. pg. 89).

1. Bob chooses an integer b, computes the point $B = bP \in \mathcal{C}_q$ and sends B to Alice (B is public).
2. Alice selects a message to encode, say a point $M \in \mathcal{C}_q$, and a random integer k. She computes

   $$A_1 = kP \qquad A_2 = M + kB$$

   and sends these points to Bob.
3. Bob recovers the message M by computing

   $$A_2 - bA_1 = M + kB - bkP = M + k(B - bP) = M$$

This is the *ElGamal* cryptosystem for elliptic curves. Its security is based on the fact that two calculations have very different levels of difficulty:

1. It is very quick to compute addition on elliptic curves ($A_1 = kP$, etc.) by computer.
2. It is very hard, given the public information P and B, to efficiently compute Bob's secret key b, without which the message cannot be decoded.

This second is the *elliptic curve discrete logarithm problem.* It is believed to be harder to solve (in terms of computing time) than the standard discrete logarithm problem.[7] This fact allows one to obtain a similar level of security using key sizes which are significantly smaller than those used for RSA, with an appreciable increase in computing speed. If you're into computing, try to write a program implementing this: remember that straightforward expressions exist for addition on elliptic curves (e.g. Exercise 7.3.5).

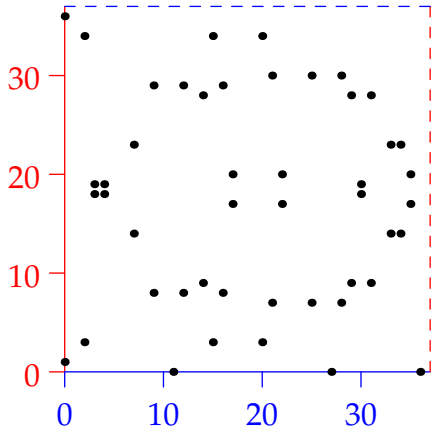
[7] We discussed the 'standard' DLP when considering primitive roots: compute b given $P, B \in \mathbb{Z}_q$ satisfying $B = P^b$.

Modularity Patterns

We finish by considering the number of points N_p on $\mathcal{C}_p$, and the defect $a_p = p - N_p$. There are many patterns here, though they are typically much more complicated than those for conics.

Examples 7.34. 1. The curve $y^2 = x^3 + 1$ has discriminant $\Delta = -27$, whence the bad primes are $p = 2$ and 3. Here are the numbers of points and their defects for several small primes:

p	Solutions (x, y)	N_p	a_p
2	$(0,1), (1,0)$	2	0
3	$(0,\pm1), (2,0)$	3	0
5	$(0,\pm1), (2,\pm2), (4,0)$	5	0
7		11	-4
11		11	0
13		11	2
17		17	0
19		27	8
23		23	0
29		0	0
31		35	-4
37		47	-10



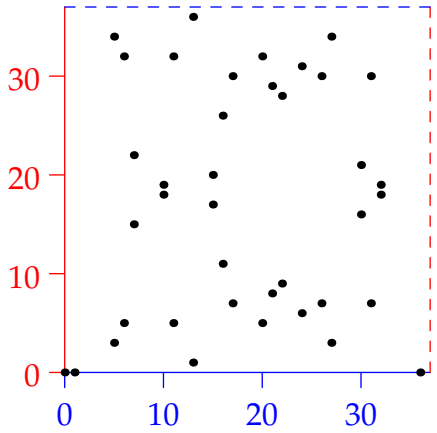
$y^2 = x^3 + 1$ modulo 37

An obvious pattern is that (for good primes) $a_p = 0$ unless $p \equiv 1$ modulo 6. This is tricky to prove, though it is related to the amazing fact that a_p is the p^{th} coefficient of the series

$$q \prod_{n=1}^{\infty} (1 - q^{6n})^4 = q - 4q^7 + 2q^{13} + 8q^{19} - 5q^{25} - 4q^{31} - 10q^{37} + \cdots$$

2. The curve $y^2 = x^3 - x$ has discriminant $\Delta = 4$ and the only bad prime is $p = 2$.

p	Solutions (x, y)	N_p	a_p
2	$(0,0), (1,0)$	2	0
3	$(0,0), (1,0), (2,0)$	3	0
5	$(0,0), (\pm1,0), (2,\pm2), (3,\pm2)$	7	-2
7	$(0,0), (\pm1,0), (4,\pm2), (5,\pm1)$	7	0
11		11	0
13		7	6
17		15	2
19		19	0
23		23	0
29		39	-10
31		31	0
37		39	-2



$y^2 = x^3 - x$ modulo 37

This time it appears that $a_p = 0$ unless $p \equiv 1$ modulo 4, which is again related to a special series

$$q \prod_{n=1}^{\infty} (1 - q^{4n})^2 (1 - q^{8n})^2 = q - 2q^5 - 3q^9 + 6q^{13} + 2q^{17} - q^{25} - 10q^{29} - 2q^{37} + 10q^{41} + \cdots$$

If you have currently have any inkling why these patterns exist, or why such series should have anything to do with the original elliptic curve, then your mathematical intuition is of genius calibre!

One can play the above game with any elliptic curve, though finding such a compact way of writing the series as an infinite product is not always possible. Since the elliptic curve determines only the *prime* coefficients a_p, we need a general method for constructing the entire series: here it is.

1. For an elliptic curve in 'minimal Weierstraß form'[8] compute the discriminant Δ and the p-defects a_p for each prime.

2. Whenever n is not prime, define the rest of the sequence by:

$$\begin{cases} a_1 = 1 & \text{if } n = 1 \\ a_{p^k} = (a_p)^k & \text{if } p \text{ is a prime dividing } \Delta \\ a_{p^{k+1}} = a_p a_{p^k} - p a_{p^{k-1}} & \text{if } p \text{ is a prime and } p \nmid \Delta \\ a_{mn} = a_m a_n & \text{if } \gcd(m,n) = 1 \end{cases}$$

3. Define a Fourier series over the complex numbers

$$f(z) = \sum_{n=1}^{\infty} a_n q^n \qquad \text{where} \quad q = e^{2\pi i z}$$

The Modularity Theorem states that this series is a *modular form*, meaning that it transforms in a special way under the action of a particular subgroup of the modular group. Specifically: there exists a positive integer $N \mid \Delta$ such that for all complex numbers $z = x + iy$ with $y > 0$, and for all $a, b, c, d \in \mathbb{Z}$ such that $ad - bc = 1$ and $N \mid c$ we have

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z)$$

This is part of what Andrew Wiles proved for all *semistable* elliptic curves (roughly curves whose reductions modulo bad primes only have double points and not cusps), which in turn was enough to establish Fermat's Last Theorem.

[8] Every elliptic curve over $\mathbb{Q}$ is 'birationally equivalent' to some curve of the form $y^2 + \alpha xy + \beta y = x^3 + \gamma x^2 + \delta x + \epsilon$ for which a suitably defined discriminant is minimal. Every elliptic curve of the form $y^2 = x^3 + ax^2 + bx + c$ considered in these notes is already minimal. The correct discriminant for these curves is 16 times ours, thus forcing $2 \mid \Delta$ for any canonical form elliptic curve.

Exercises 1. Factorize $y^2 = x^3 - 2x + 1 = (x-1)(x^2+x-1)$. This holds modulo any prime p.

(a) Check that $x = 1$ is not a solution to $x^2 + x - 1 = 0$ in $\mathbb{Z}_p$ for any prime p.

(b) Prove that $x^2 + x - 1 = 0$ has a repeated root in $\mathbb{Z}_p$ if and only if $p = 5$.
(*Hint: multiply out* $(x-\mu)^2 = x^2 + x - 1$. *What conditions must* μ *satisfy?*)

(c) Find all the points on the curve modulo $p = 2, 3, 5$ and 7. Show that there are no singular points when $p = 3$ and 7, and find the singularities when $p = 2$ and 5.

This should convince you that the only bad primes for the curve are $p = 2$ *and 5.*

2. (a) Prove that if either a or b is non-zero modulo p, then $ax + by = c$ has $N_p = p$ (equiv. $a_p = 0$).

(b) What happens if $a = b = 0 \in \mathbb{Z}_p$?

(c) Suppose n is composist and that a, b are units in $\mathbb{Z}_n$. What is N_n?

(d) How many points are there on the curve $2x + 4y = 6$ over $\mathbb{Z}_{10}$ and how does this relate to the modulus?

3. For each elliptic curve, find the discriminant and the primes of bad reduction. When $p \neq 2$ is a bad prime find the singular point(s) on $\mathcal{C}_p$ and identify their type (double point or cusp).

(a) $y^2 = x^3 - 2x + 1$

(b) $y^2 = x^3 - 5x + 4$

4. The curve $\mathcal{C} : 5x^2 - y^2 + x + 1 = 0$ contains the point $(0, 1)$.

(a) Find all the points on $\mathcal{C}_2$, the reduction of the curve modulo 2.

(b) By considering $9^2(x+2)^2$ modulo 19, obtain a factorization of $\mathcal{C}$ into two lines. Compute the number of points N_{19} on $\mathcal{C}_{19}$ and the defect $a_{19} = 19 - N_{19}$.

(c) Suppose that $p \neq 2, 19$ is prime. Prove that the 'line' $y = mx + 1$ intersects $\mathcal{C}_p$ at a second point if and only if $m^2 \neq 5 \in \mathbb{Z}_p$. What are the co-ordinates of the new point?

(d) If $p \neq 2, 5, 19$ is prime, prove that the defect is given by

$$a_p = \left(\frac{5}{p}\right) = \begin{cases} 1 & \text{if 5 is a quadratic residue modulo } p \\ -1 & \text{if 5 is a quadratic non-residue modulo } p \end{cases}$$

What happens if $p = 5$?