ICAMMP 2004
4-7 January 2005
SUST, Sylhet, Bangladesh

Cubic curves: a short survey

Balázs Szendrői
Department of Mathematics, University of Utrecht, The Netherlands
szendroi@math.uu.nl http://www.math.uu.nl/people/szendroi

Abstract: *This article discusses cubic curves, some of the ways they arise, and some of their applications in diverse fields such as cryptography and string theory.*

Key-Words: cubic curve, elliptic curve, Weierstrass equation

1 Introduction

The purpose of this article is to discuss, in a light-hearted way, various occurences of the cubic Weierstrass equation

$$y^2 = x^3 + ax + b \tag{1}$$

in and out of mathematics. I begin by explaining how this equation defines a curve in the plane, and draw some pictures. Next I recall one of the earliest historic occurences of the cubic equation (1), in the computation of the arc length of the ellipse. In Section 4, I will discuss the connection between cubic curves over the complex numbers and the torus. Section 5 will introduce a famous attribute of cubic curves, their group law. This will lead in Section 6 to a discussion of the use of cubic curves in cryptography. Finally in Section 7 I will sketch how cubic elliptic curves turn up in superstring theory.

The length of this article does certainly not permit me to do full justice to the importance of cubic curves. Even on the level of New York Times headlines, there are at least two glaring omissions: the relationship between cubic curves and Fermat's Last Theorem; and the Birch–Swinnerton-Dyer conjecture, one of the Clay Institute's $10 lakh questions. Their story would take me too far afield; [1] and [2] both make good read following on from this article. See also [3] for further information and more general references on the subject, which in turn include proofs of all statements left unproved in this article.

This article is a write-up of my talk at the International Conference in Applied Mathematics and Theoretical Physics at Shahjalal University of Science and Technology, Sylhet, Bangladesh. I wish to thank all my hosts, especially Professor Anwar Hossain of Dhaka University, for their wonderful hospitality.

2 The cubic curve

Recall our basic equation

$$y^2 = x^3 + ax + b.$$

In this equation, we usually think of a, b as fixed constants, and x, y as variables. Choosing a field K, and taking values for $a, b \in K$, the pairs (x, y) satisfying equation (1) form a subset of the Cartesian product $K \times K$.

Now think of the product $K \times K$ geometrically as the *plane* over the field K. The set of points (x, y) satisfying the cubic equation forms a subset of the plane; this is the *cubic curve*

$$C = \{(x, y) : y^2 = x^3 + ax + b\} \subset K^2.$$

Figure 1 shows pictures of some cubic curves in the plane over the real numbers $K = \mathbb{R}$.

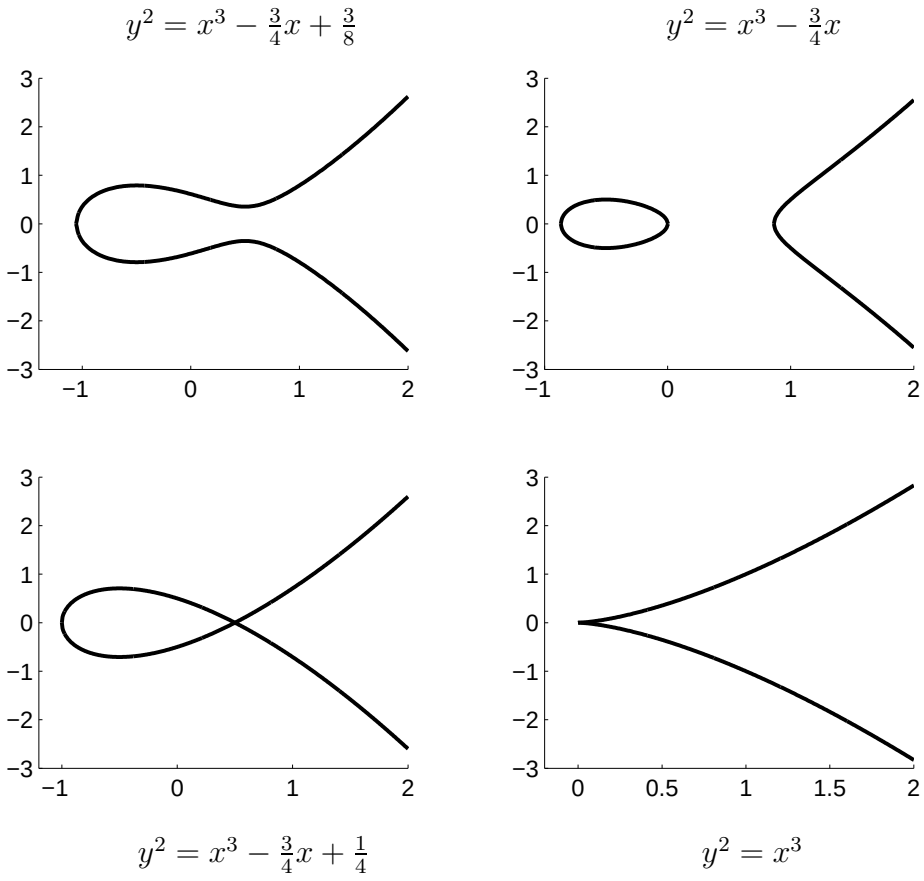

Figure 1: Some cubic curves over the real numbers

Properties of the curve C depend in a basic way on the cubic polynomial $p(x) = x^3 + ax + b$. Extending the field K if necessary, $p(x)$ factors into a product of linear factors as

$$p(x) = (x - \lambda_1)(x - \lambda_2)(x - \lambda_3).$$

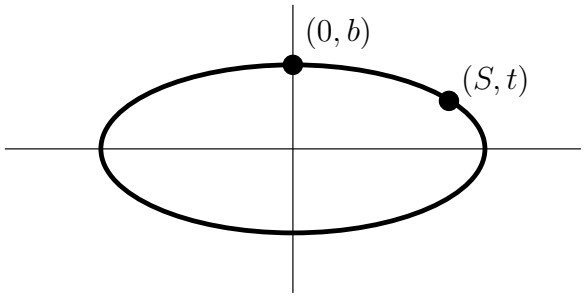

Figure 2: The ellipse in the plane

- C is a *smooth curve*, if the three roots λ_i are all different.
- C is a *nodal curve*, if two of the λ_i coincide and the third one is different.
- C is a *cuspidal curve*, if all three roots λ_i are identical.

As one can check easily (Exercise!), the first two curves in Figure 1 are smooth, the third one is nodal, whereas the last one is cuspidal. The latter curves have special points: the third curve intersects itself at the *node* $(\frac{1}{2}, 0)$, whereas the last curve contains the *cusp* $(0, 0)$, where it does not have a well-defined tangent direction.

3 From elliptic integrals to cubic curves

Consider the standard ellipse in the real (s, t) plane with equation

$$\frac{s^2}{a^2} + \frac{t^2}{b^2} = 1 \tag{2}$$

as pictured on Figure . The arc length along the ellipse, from the point $(0, b)$ to some variable point (S, t), is given by

$$\begin{aligned} L(t) &= \int_0^S \sqrt{1 + \left(\frac{\partial}{\partial s} t(s)\right)^2}\, ds \\ &= \int_0^S \left(1 + \frac{b^2 s^2}{a^4\left(1 - \frac{s^2}{a^2}\right)}\right)^{\frac{1}{2}} ds \\ &= \int_0^{S/a} \left(\frac{1 - e^2 u^2}{1 - u^2}\right)^{\frac{1}{2}} du, \end{aligned}$$

where I have set $u = s/a$, and $e^2 = 1 - b^2/a^2$.

If $e = 0$, then this integral can be evaluated explicitly (Exercise!). $e = 0$ is equivalent to $a = b$, when the ellipse simplifies to a circle. On the other hand, as possibly noticed first by Legendre, if $e \neq 0$, then the integral cannot be evaluated in terms of elementary functions. The best one can do is find a geometric object

on which this integral naturally lives. If v denotes the integrand, then the following relation holds between u and v:

$$u^2v^2 - e^2u^2 - v^2 + 1 = 0. \tag{3}$$

This is not yet a cubic equation; rather, this is a quatric equation in the variables (u, v) of a rather special form. To transform this quartic, introduce new variables (x, y) by

$$x = \frac{2}{1-u}, \quad y = \frac{2}{\sqrt{1-e^2}}\frac{v(1+u)}{1-u}.$$

Such a change of variables is called a *rational map*, since the expressions for (x, y) are rational functions of (u, v). This rational map is invertible:

$$u = \frac{x-2}{x}, \quad v = \frac{\sqrt{1-e^2}}{2}\frac{y}{x-1}.$$

Also, as a patient calculation shows (Exercise!), the change of variables $(u, v) \to (x, y)$ transforms equation (3) into the following one:

$$y^2 = x^3 - \frac{5e^2-1}{e^2-1}x^2 + 8\frac{e^2}{e^2-1}x - 4\frac{e^2}{e^2-1}. \tag{4}$$

It is now a simple matter to make a linear change of variables in x (Exercise!) to bring equation (4) into the form (1).

To summarize: given an ellipse (2), its arc length can be expressed as an integral $\int v\,du$, where (u, v) are related, after an invertible rational change of variables, by a standard cubic equation (1). It is easy to check (Exercise!) that $e \neq 0$ corresponds to a smooth cubic curve, whereas $e = 0$ gives a nodal curve. I can thus say that integrals such as that computing the arc length of an ellipse with $e \neq 0$, henceforth called *elliptic integrals*, are intimately related to the geometry of smooth cubic curves. Smooth cubic curves are also called *elliptic curves* accordingly.

4 Tori and cubic curves

I now describe a completely different route to cubic curves, involving the complex number field $\mathbb{C}$. Start with the square lattice $\mathbb{Z}^2$ contained in the real plane $\mathbb{R}^2$. The lattice can be thought of as a group of translations acting on $\mathbb{R}^2$, and correspondingly there is a topological quotient space $\mathbb{R}^2/\mathbb{Z}^2$. This quotient can be obtained from one fundamental square of the lattice with its sides glued together appropriately; as Figure shows, the result is a topological *torus*.

Now think of the real plane $\mathbb{R}^2$ as the space $\mathbb{C}$ of complex numbers, coordinatized by $z \mapsto (\operatorname{Re} z, \operatorname{Im} z)$. As it turns out, $\mathbb{C}$ admits different embedded lattices (up to analytic equivalence): take the lattice Λ_τ generated by 1 and a complex number τ, lying in the upper half plane, as on Figure . The *analytic torus* is the quotient $\mathbb{C}/\Lambda_\tau$; topologically, all these spaces are the same (homeomorphic), but their complex analytic structure varies with τ.

Next, consider the function defined on $\mathbb{C} \setminus \Lambda_\tau$ by a (convergent) infinite sum

$$\wp_\tau(z) = \frac{1}{z^2} + \sum_{w \in \Lambda_\tau \setminus 0} \left(\frac{1}{(z-w)^2} - \frac{1}{w^2} \right).$$

After its inventor, this function is called the *Weierstrass $\wp$-function.* It can be proved to be an analytic function on the complement $\mathbb{C} \setminus \Lambda_\tau$ of the lattice Λ_τ.

Theorem 1 *The function $\wp_\tau$ and its derivative $\wp'_\tau$ satisfy the following two fundamental identities.*

- Periodicity*: for all $w \in \Lambda_\tau$,*

$$\begin{aligned} \wp_\tau(z) &= \wp_\tau(z+w) \\ \wp'_\tau(z) &= \wp'_\tau(z+w). \end{aligned} \tag{5}$$

- The Weierstrass cubic relation*: there exist complex numbers a_τ, b_τ, so that*

$$\wp'_\tau(z)^2 = 4\wp_\tau(z)^3 + a_\tau \wp_\tau(z) + b_\tau, \tag{6}$$

with the cubic $4x^3 + a_\tau x + b_\tau$ having different roots over $\mathbb{C}$.

To translate these identities into a geometric statement, consider the map

$$\begin{array}{ccc} \mathbb{C} \setminus \Lambda_\tau & \to & \mathbb{C} \times \mathbb{C} \\ z & \mapsto & (\wp_\tau(z), \wp'_\tau(z)) \, . \end{array}$$

Because of the periodicity relations (5), this map takes points of $\mathbb{C}$, identified under the action of Λ_τ, to the same point of $\mathbb{C} \times \mathbb{C}$. Hence I can regard this map as

$$\begin{array}{ccc} (\mathbb{C} \setminus \Lambda_\tau)/\Lambda_\tau & \to & \mathbb{C} \times \mathbb{C} \\ z & \mapsto & (\wp_\tau(z), \wp'_\tau(z)) \, . \end{array}$$

On the other hand, by Weierstrass' relation (6), the image of this map is contained in the smooth cubic curve given by that relation, so there is a map

$$(\mathbb{C} \setminus \Lambda_\tau)/\Lambda_\tau \longrightarrow C_\tau = \{y^2 = 4x^3 + a_\tau x + b_\tau\} \subset \mathbb{C} \times \mathbb{C}.$$

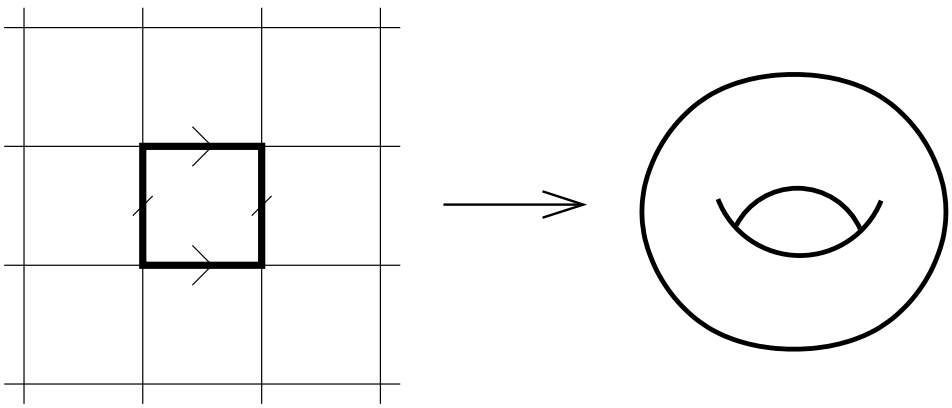

Figure 3: The quotient $\mathbb{R}^2/\mathbb{Z}^2$ is a torus

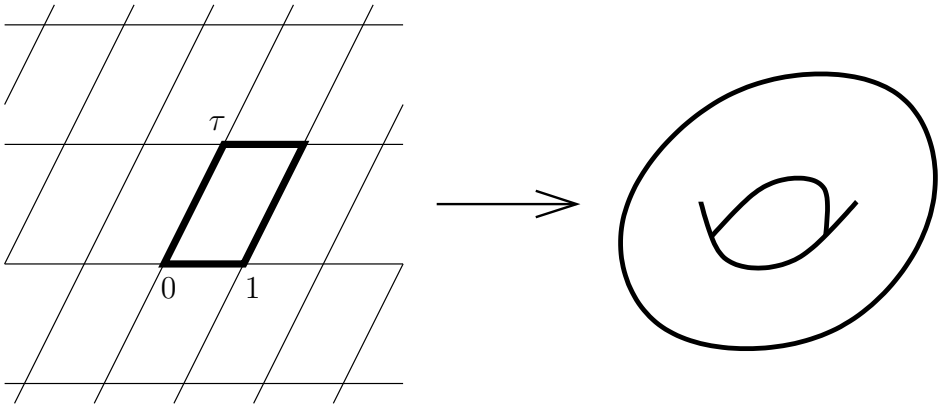

Figure 4: The lattice Λ_τ and the quotient $\mathbb{C}/\Lambda_\tau$

One of the important results of this theory is that this map is an *isomorphism* of complex analytic spaces. On the right hand side is a cubic curve, given by an equation of the type (1). On the left hand side is "most" of the torus $\mathbb{C}/\Lambda_\tau$; as the $\wp$-function is not defined at the lattice points, one point is missing. To compensate for that, append the cubic curve by a special point ∞ (the formal way to do that would involve projective coordinates, a subject I do not go into here) to get an analytic isomorphism between our torus and a smooth cubic curve:

$$\mathbb{C}/\Lambda_\tau \xrightarrow{\sim} \{y^2 = 4x^3 + a_\tau x + b_\tau\} \cup \{\infty\}.$$

This cubic curve lives in the *complex 2-plane* $\mathbb{C} \times \mathbb{C}$, which is a *four-dimensional* real space; one complex equation leads to a *two-dimensional* real space, in other words our torus!

The converse result is also true (though a little more difficult to prove): all smooth cubic curves come from tori. I summarize the discussion of this section in the following result, including also a description of the geometry of non-smooth cubic curves (extended by a point ∞).

Theorem 2 *1. The analytic torus $\mathbb{C}/\Lambda_\tau$ is isomorphic to the smooth cubic curve $C_\tau \cup \{\infty\}$, given by equation (6). Conversely, every smooth cubic curve over the complex number field $\mathbb{C}$ is isomorphic to an analytic torus $\mathbb{C}/\Lambda_\tau$.*

2. A nodal curve over $\mathbb{C}$ is a geometrically a degenerate pinched torus*, pictured on Figure 5; this shape is really just a sphere, with two different points identified at the node.*

3. The geometry of a cuspidal curve over $\mathbb{C}$ is that of a sphere*, marked by a special point, the position of the cusp.*

With this geometric description in hand, let us return to the story of the elliptic integral $\int v\,du$, discussed in the previous section. For non-zero parameter e, the complex cubic curve (4) is smooth, hence it is a torus. If however $e = 0$, corresponding to arc length on the circle, the cubic equation (4) becomes *nodal* (Exercise!). The

computation of the arc length of the circle happens on the sphere corresponding to this nodal curve; the answer of such an integral is readily computed in terms of trigonometric functions. In other words, the geometry of cubic curves is complicated as long as the corresponding geometric shape is a torus; in the special case when it reduces to a sphere, the complications disappear.

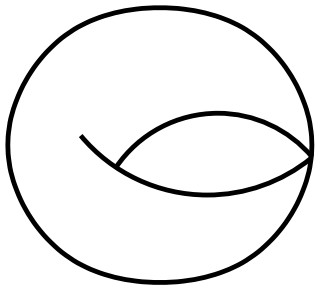

Figure 5: A degenerate torus

5 The group law on a cubic curve

Our next topic is a remarkable algebraic construction on a smooth cubic curve over a field K, which exploits the fact that the degree of its defining equation is precisely *three*. Take two points $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ on a cubic curve C in the (x, y) plane, and draw the line PQ. This line has equation $y = mz + n$, and substituting into (1) results in a *cubic* equation for the x-coordinates of the intersection points (Exercise!). Two of these are x_P and x_Q, so there is a third solution, leading to a third intersection point R; compare Figure . This construction can be extended to special cases: if $P = Q$, draw the tangent to the curve; if P, Q have the same x-coordinate, so that the PQ line has equation $x = c$, then let $R = \infty$ be the extra point.

With these rules, I associate to every pair of points P, Q on $C \cup \{\infty\}$ a third point R. This can be used to define an addition $\boxplus$ on points of the set $C \cup \{\infty\}$ by declaring

$$P \boxplus Q \boxplus R = 0$$

whenever R is the point associated to (P, Q) as above.

Theorem 3 *The operation $\boxplus$ on points of $C \cup \{\infty\}$ is associative, commutative, has an identity ∞, and inverses. Thus the set of points of a smooth cubic curve over a field K forms a commutative group under the operation $\boxplus$.*

One significant point regarding Theorem 3 is that holds over any field K, and thus results a good supply of commutative groups defined over any field. Since nodal and cuspidal curves are also defined by a cubic equation, much of the discussion extends to them also; note though, that the node, respectively the cusp has to be removed (Exercise: Why?). The resulting groups however turn out to be well known.

Theorem 4 *The set of non-special points of a nodal, respectively cuspidal curve over a field K form a commutative group. The group of points of a nodal curve is*

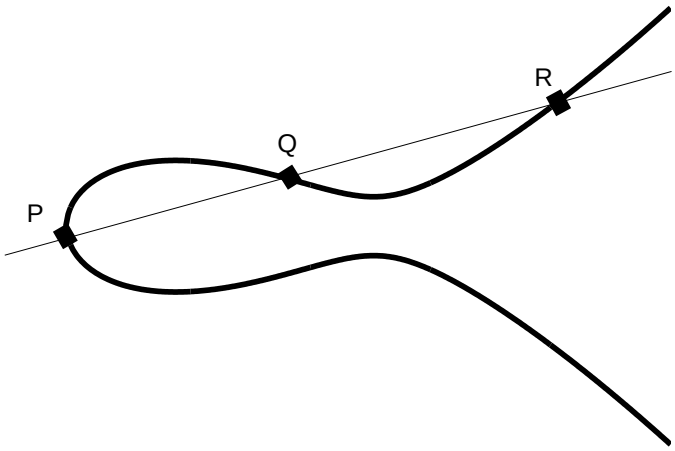

Figure 6: The construction of the group law

isomorphic to the multiplicative group $(K^*, \cdot)$ *of the field. The group of points of a cuspidal curve is isomorphic to the additive group* $(K, +)$.

Specializing to the case $K = \mathbb{C}$, recall that in the previous section I set up for every smooth cubic curve C_τ an isomorphism with a torus $\mathbb{C}/\Lambda_\tau$. Under this isomorphism, the group operation $\boxplus$ corresponds to the obvious abelian group operation on the set of points of $\mathbb{C}/\Lambda_\tau$, coming from the addition of complex numbers.

6 Cubic curves in cryptography

The group law means that there is *arithmetic* on points of a cubic curve: I can add and subtract, and thus store information in points of said curve. Nodal or cuspidal curves do not yield anything new by Theorem 4, but for smooth curves the situation is different. In was proposed in the 1980s that smooth cubic curves could be used for a new kind of cryptographic scheme.

In order to be able to implement any such scheme on a computer, restrict to a finite field $K = \mathbb{F}_p$, the arithmetic of integers modulo a prime number p, or more generally to a finite field $\mathbb{F}_q$ with $q = p^k$ elements. A smooth cubic curve C over such a field $\mathbb{F}_q$ has finitely many points (Exercise!), and the set of these points is equipped with a group structure. In particular, given a point $P \in C$, it makes sense to form $P \boxplus P$, $P \boxplus P \boxplus P$, and more generally for any integer n, the point $nP = P \boxplus \ldots \boxplus P \in C$. *Elliptic curve cryptography* is based on the

Elliptic curve logarithm problem: Given points P, Q on a cubic elliptic curve, determine the smallest integer n (if it exists) such that

$$Q = nP.$$

This problem is believed (though not proven!) to be theoretically more difficult than the corresponding problem in the multiplicative group of a finit field $\mathbb{F}_q$, on which most current cryptosystems are based.

In a cryptographic context, one fixes a smooth cubic curve together with a point (C, P), and encodes information in its multiples nP. For practical applications, the point P has to have large order, so that there is room for ample information storage; it is therefore an important issue to find interesting elliptic curves over finite fields with points of large order.

7 Cubic curves in superstring theory

Our final theme will take us on a short tour into the realm of theoretical physics. Superstring theory, originally proposed as a (failed) account of the strong force in particle physics, has emerged in the 1980s as a candidate for a *universal theory of physics*, bridging the gap between "the theory of the small", quantum mechanics, and "the theory of the large", general relativity. It is based on the idea that the basic building blocks of the universe are not point particles, but one-dimensional wiggling strings, along with other, higher dimensional geometric objects. The background space-time $\mathcal{M}$ of superstring theory is strongly constrained by various consistency requirements; in particular, one possibility is that it has the form $\mathcal{M} = \mathbb{R}^{1,d} \times X$. Here $\mathbb{R}^{1,d}$ is our well known and loved Minkowski space, whereas X is a compact geometric space with a complex structure and a special (Ricci-flat Kähler) metric.

Various (hard) theorems of differential geometry ensure that there is in fact a plentiful supply of suitable geometric spaces, the so-called *Calabi–Yau manifolds*. Among these, there is one class of spaces where one can dispense with the hard theorems; our tori $X = \mathbb{C}/\Lambda_\tau$ can be used as suitable string backgrounds. By construction, the torus X has a complex analytic structure, and it has a nice, completely flat metric $\omega dx \wedge dy$ derived from the identification of $\mathbb{C}$ with flat space $\mathbb{R}^2$, the positive real number ω determining the overall size of the torus. The space of metrized complex tori is therefore parametrized by the pair (τ, ω), where the complex parameter τ is in the upper half plane of $\mathbb{C}$, whereas the metric parameter $\omega > 0$ is real.

One powerful idea emerging in the field of theoretical physics is that of *dualities*, meaning the complete physical equivalence of different theories constructed from different geometric backgrounds. One duality that applies in the present context is *mirror symmetry*, identifying superstring theories defined by two different Calabi–Yau spaces of the same dimension, interchanging metric and complex analytic data.

The only one-dimensional Calabi–Yau spaces are our tori $(\mathbb{C}/\Lambda_\tau, \omega dx \wedge dy)$. Thus mirror symmetry in this case says that there is a self-map of the parameter space of metric complex tori, interchanging metric and complex data. This however presents a puzzle: the space of our tori is the space of pairs (τ, ω), which is notably non-symmetric: the complex data depends on a point of the upper half plane, whereas the metric data depends on a positive real number.

The resolution of the puzzle comes from physics, where it was realized that a superstring theory depends on an additional metric-type parameter, the so-called *B-field* $B \in H^2(X, \mathbb{R})$. In the case of a torus, the latter cohomology space is isomorphic to $\mathbb{R}$. Moreover, the two parameters B, ω can be combined into the complex parameter $B + i\omega$, which now conveniently lies in the upper half plane. Mirror symmetry for elliptic curves now says the following.

Physics Theorem 5 *Physics on the elliptic curve* $(\mathbb{C}/\Lambda_\tau, B + i\omega)$*, with complex parameter* τ *and metric parameters* $B+i\omega$*, is identical to physics on the elliptic curve* $(\mathbb{C}/\Lambda_{B+i\omega}, \tau)$*, with complex parameter* $B + i\omega$ *and metric parameters* $\mathrm{Re}\,\tau + i\,\mathrm{Im}\,\tau$*. The identification interchanges metric and complex analytic structures on the two elliptic curves in a non-trivial way.*

This statement, in its different physical and mathematical manifestations, has been an interesting testing ground for all sorts of conjectures in the field of mirror symmetry. The study of mirror symmetry for cubic elliptic curves has also lead to powerful ideas which could then be applied in the much more complicated case of higher dimensional Calabi–Yau manifolds, notably quartic surfaces and quintic threefolds. I conclude the tour around the remarkable story of cubic elliptic curves on this happy and positive note.

References

[1] Clay Mathematical Institute: Birch–Swinnerton-Dyer conjecture
`http://www.claymath.org/millennium/Birch_and_Swinnerton-Dyer_Conjecture/`

[2] G. Faltings: The proof of Fermat's Last Theorem by R. Taylor and A. Wiles, Notices of the American Mathematical Society, July 1995
`http://www.ams.org/notices/1995007/faltings.pdf`

[3] mathworld by Wolfram Research: Elliptic curve
`http://mathworld.wolfram.com/EllipticCurve.html`