Topological Interpretation of Elliptic Curves

Topological Interpretation of Finite Cyclic Groups

Avneet Singh
Interplanetary Company UG
sshmatrix@proton.me

ABSTRACT

Elliptic curves form the backbone of most cryptography today and are expected to feature in the post-quantum world through Zero-Knowlege algorithms. Most literature on Elliptic curves starts from the definition of a cubic symmetric polynomial and builds upon the group theoretic interpretation of finite fields. This may be sufficient to grasp the mere necessary details for implementing Elliptic curves in practise, but not necessarily ideal if one wants to understand why Elliptic curve geometry is unique and useful. To most readers, it can appear that Elliptic curves were drawn from a magic hat. In paper will illustrate that this is clearly not the case and there exists a very legitimate reason for how and why the humanity ended up using Elliptic curves for cryptography.

INTRODUCTION

This paper is intended as a standalone appendix to the main paper titled 'Intuitive Interpretation of Non-Interactive Zero-Knowledge Cryptography' [1]. The main paper deals with pretty much the entire field of cryptography starting from basic RSA to advanced Zero-Knowledge algorithms. In interest of conciseness, several important aspects of Elliptic curves couldn't be discussed in detail in the main paper; this document is an attempt to address those shortcomings for the overtly inclined reader. We'll also touch some peripheral topics that have tied number theory and elliptic curves at the hip. Having said that, we'll follow the same theme as the main paper and avoid jargon like plague.

ARC-LENGTH OF AN ELLIPSE

The hint is in the name. Elliptic curves are related to ellipses and their discovery was 'accidental' so to speak; most of the initial work toward elliptic curves was done by Newton, Legendre, Fermat, Euler, Abel and Jacobi. In simple terms, an elliptic curve is a paramterisation of an ellipse's circumference or arc length and it easily derivable. To begin with, consider an ellipse described by its semi-major and -minor axis lengths of α and γ (=1), i.e. eccentricity $e^2 = 1 - 1/\alpha^2$. Such an ellipse is described by: $(1 - e^2)x^2 + y^2 = 1$. We can simplify it further by safely[1] replacing the constant $1 - e^2$ with k^2, yielding $y^2 = 1 - k^2x^2$. Let's try to calculate the circumference $C(k)$ of this ellipse; this will be equivalent to integrating $(\delta x^2 + \delta y^2)^{1/2}$ along the curve in one of the four cartesian quarters and then multiplying it by 4. This results in

$$C(e) = 4 \int (\delta x^2 + \delta y^2)^{1/2}, \text{ for } x = [0, \alpha], \tag{1}$$

---

[1] such that the values of both e and k lie between 0 and 1, aka [0, 1]

$$C(e) \;=\; 4\int_0^{\alpha}(\delta x^2 \,+\, \delta y^2)^{1/2} \;=\; 4\int_0^{\alpha}\left[\frac{1\,-\,e^2x^2}{1\,-\,x^2}\right]^{1/2}\delta x. \tag{2}$$

This is already the **elliptic integral** of the second kind[2] and it is particularly difficult to evaluate analytically. When e = 1 (i.e. straight lines; k = 0), it is straightforward and evaluates to 4α as expected. When e = 0 (i.e. circles; k = 1), it is also relatively straightforward and evaluates naturally to

$$C(e)_{k=1} \;=\; 4\int_0^{\alpha}\frac{1}{(1\,-\,x^2)^{1/2}}\,\delta x \;=\; 4\left.\vphantom{\int}\right|_0^{\alpha}\sin^{-1}(x) \;=\; 2\pi\alpha. \tag{2a}$$

In conjunction with the well-known trigonometric parameterisation of ellipse with $(x,\,y)\to(\alpha\sin\theta,\,\gamma\cos\theta)$, the general elliptic intergal reduces to,

$$C(e) \;=\; 4\int_0^{\pi/2}(1\,-\,e^2\sin^2\theta)^{1/2}\,\delta\theta. \tag{3}$$

In strict sense, the inverse of the integrand in the elliptic integral (2)–(3) is already the elliptic curve, but it is not straightforward to interpret in this form. In order to arrive at a more intuitive and interpretive form of elliptic curve, we perform a parameterisation of the form $\theta^n = 1 - e^2x^2$. Let's consider in particular the parametrisation for n = 1, such that $\theta = 1 - e^2x^2$. This leads to

$$C(k) \;=\; \frac{1}{2}\int_{k^2}^{1}\frac{\theta}{[\theta(\theta\,-\,k^2)(\theta\,-\,1)]^{1/2}}\,\delta\theta \;=\; -\frac{\mathbf{i}}{2}\int_{k^2}^{1}\frac{\theta}{E(\theta)}\,\delta\theta, \tag{4}$$

where $\mathbf{i}$ is the usual complex unit, i.e. $\mathbf{i} = \sqrt{-1}$. The 3-order polynomial term in the denominator is of the form
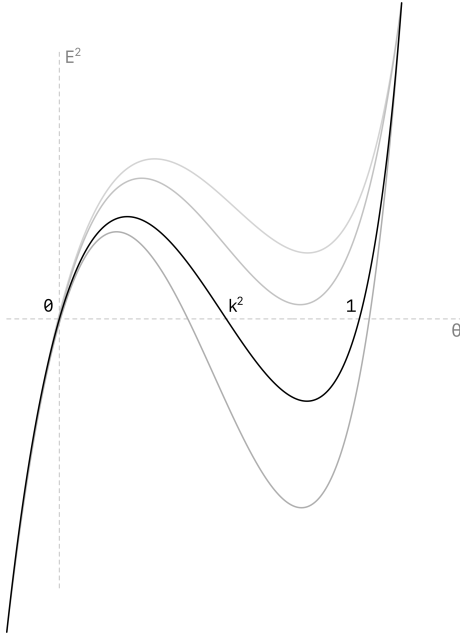

Figure 1:   Polynomial $E^2(\theta)$ with one and three real root(s) in $\mathbb{R}$ space
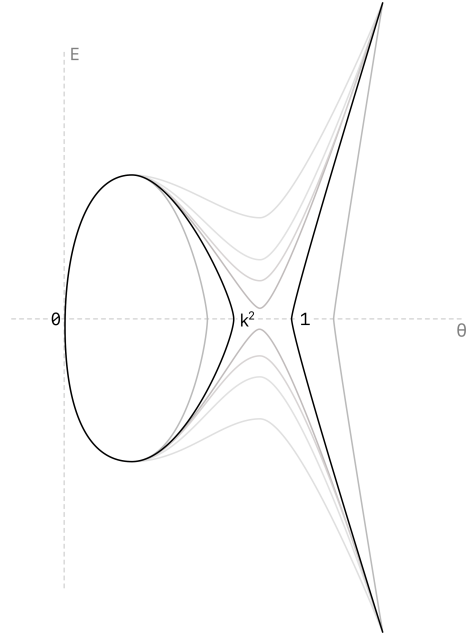
Figure 2:   Family of curves $E(\theta)$ with one and three real root(s) in $\mathbb{R}$ space

---

[2]'second kind' due to the presence of θ in the numerator in its parameterised form; 'first kind' has 1 in the numerator

$$E^2(\theta) = \theta^3 + r \cdot \theta^2 + p \cdot \theta + q,$$

which is precisely the more familiar and recognisable form of the **elliptic curve**! Note that the form of $E(\theta)$ in the above example asserts three roots of $E(\theta)$, i.e. 0, k^2, 1. This is however not true for generic elliptic curves which may have only one real root depending on the values of p, q and r, all of which are functions of k in return. In our particular example, $r = -(1 + k^2)$, $p = k^2$ and $q = 0$. Some readers may feel cheated at this point since we have just said that an elliptic curve is simply the square root of a cubic polynomial! This is indeed correct but not every cubic polynomial is an elliptic curve; the coefficients of the cubic polynomial must be such that there are no repeated roots, i.e. they represent a legitimate ellipse with legal values of e^2 and k^2. Lastly, we note that the if the elliptic curve has 3 real roots, then evaluation of (4) requires integrating $E(\theta)$ in the range $(k^2, 1)$ where it is purely complex! In other words, figure 6 doesn't paint the entire picture of elliptic curves with three real roots instead of one. To fully grasp the topological intuition behind $E(\theta)$, we must include the complex plane in our graphics. We also make a note of the parameterisation $\theta = 1 - e^2x^2$ that led to the easier form of the elliptic curve in (4); what led to this choice?[Q1]

PERIODICITY

It was Abel and Jacobi who first understood that the inverse of the intergal in (2a) is periodic and more relevant than the intergal itself, i.e.

$$F(\bar{x}) = \int_0^{\bar{x}} \frac{1}{(1 - x^2)^{1/2}} \, \delta x = \sin^{-1}(\bar{x}) = \theta. \tag{5}$$

Abel and Jacobi crucially formalised the properties of inverses of trigonometric functions using definite integrals and differential & intergal calculus developed by Newton more than 150 years prior. The assertion of periodicity in Abel's works triggered several prominent number theorists to take note of elliptic integrals and their remarkable properties. For instance, the above integral becomes solvable only once we insert the parameterisation $(x, y) = [f(\theta), f'(\theta)] \to (\sin\theta, \cos\theta)$! Based on this observation, Abel and Jacobi asserted that the key to decoding elliptic intergals is to find a similar parameterisation (f, f') which reduces the elliptic intergal to an integrable form. Thus, the search had begun for such parameterisations, primarily by Jacobi and Abel.

Legendre in the past had re-arranged equation (2) in another form which turned out to be more useful for Jacobi. Legendre's form of (2) simply reads

$$C(e) = 4\alpha \int_0^1 \frac{1 - e^2x^2}{[(1 - x^2)(1 - e^2x^2)]^{1/2}} \, \delta x, \tag{6}$$

Legendre noted that integrating the elliptic integral of second kind is easier if one knows the integral evaluation of the first kind[2], i.e. the same equation as above but with the numerator set to 1,

$$C(\bar{x}) = 4\alpha \int_0^{\bar{x}} \frac{1}{[(1 - x^2)(1 - e^2x^2)]^{1/2}} \, \delta x. \tag{7}$$

---

[2]one can observe this by integrating (6) using the elementary 'by parts' formalism

In parameterised form, elliptic integral of the first kind reads,

$$C(e) \;=\; \int_0^{\pi/2} \frac{1}{(1 - e^2\sin^2\theta)^{1/2}}\, \delta\theta, \tag{8}$$

which is eerily innocent–looking and quite similar to (3) except that the integrand is inverted. This particular form of the elliptic integral of the first kind is where Jacobi made the first meaningful step toward parameterising generic elliptic curves. Using nothing more than elementary geometry and trigonometry, Jacobi found $C(\bar{x})$ to be analogous to a sine–inverse–like function[2] (denoted $sn^{-1}x$, $cn^{-1}x$, $dn^{-1}x$ etc), and found out that it did in fact have properties similar to a sinusoid such as periodicity. In fact, $C(\bar{x})$ is a bi–periodic function with complex periods! In simpler words, equation (7) is doubly periodic with both periods ($\hat{\omega}_1, \hat{\omega}_2 \in \mathbb{C}$) being complex numbers, such that

$$C(\bar{x}) \;=\; C(\bar{x} + n_1 \cdot \hat{\omega}_1) \;=\; C(\bar{x} + n_2 \cdot \hat{\omega}_2), \tag{9}$$

where $\hat{\omega}_1$ and $\hat{\omega}_2$ must necessarily be linearly–independent in $\mathbb{R}$, aka they are not mutually related by $\hat{\omega}_1 \neq \mathbb{R} \cdot \hat{\omega}_2$. This however wasn't the most important assertion; Jacobi further found that no single–variable function could possibly have more than two independent complex periods, and all except elliptic integrals of the first and second (and third) kind have at most one independent period only! This was arguably the first solid proof that elliptic integrals were a special class of functions.

COMPLEX BI–PERIODICITY

We are now at a point where we cannot avoid complex numbers any further. We faced them while trying to evaluate the elliptic integral when $E(\theta)$ had three real roots, and now we are facing them yet again when trying to retrace Jacobi's steps. In order to understand the topology of Jacobi's trigonometric functions, we must first introduce complex numbers and particularly n–dimensional lattices in complex space, both of which are necessary to understand bi–periodic functions with complex periods.

Periodicity in real space $\mathbb{R}$ is simple to understand topologically. Complex numbers on the other hand behave as vectors in geometrical sense, and therefore periodicity in a vector space must be understood. For example, consider equation (8) defining periodicity in complex space with periods $\hat{\omega}_1$ and $\hat{\omega}_2$. Consider any one of the periods $\hat{\omega}_1$ to begin with and calculate a few iterations of of $\bar{x} + n_1 \cdot \hat{\omega}_1$ for a few values of $n_1 = 1, 2, ... 6$. We find that the corresponding vectors all lie on a straight line in the complex plane parallel to $\hat{\omega}_1$; these points are shown in figure 3 with their indices. In the same way, the vectors associated with the second period also generate a straight line with a different slope parallel to $\hat{\omega}_2$. The two sets of points lying on lines of unequal slope in the complex plane form a lattice, denoted by $\hat{\omega}_1 \times \hat{\omega}_2$, such that the bi–periodic function is also periodic on the lattice. The trivial extension of this is that when both periods are linearly–dependent, the lattice is in fact just a straight line. Recall that for a mono–periodic real functions such as $\sin(x)$, its parameterised form is a closed curve (a circle). In other words, a mono–periodic function parameterises to a closed curve in 2–dimensional space. In the same way, a bi–periodic function parameterises to a closed surface in 3–dimensional space. Does this mean that a bi–periodic $C(\bar{x})$ is a closed surface

---

[2]due to the said integral being literally $\sin^{-1}(\bar{x})$ for $e = 0$

in complex hyperplane[2]? Yes. How does this surface look? Figure 4 shows a visualisation of the parameterisation similar to how a circle can be obtained by parameterising a sinusoid by θ. The form of a bi-periodic C($\bar{x}$) is a complex torus, the only closed surface capable of bi-periodicity in three dimensions. It can be intuitively derived by rolling the lattice sheet into a cylinder along $\hat{\omega}_1$, and then joining the opposite ends of the rolled cylinder (by 'rolling' the cylinder second time along $\hat{\omega}_2$) to form a torus.
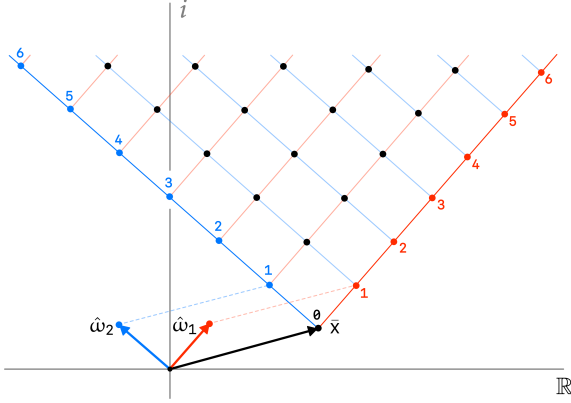
Figure 3:   Complex lattice generated by a bi-periodic function with two linearly in-dependent complex periods
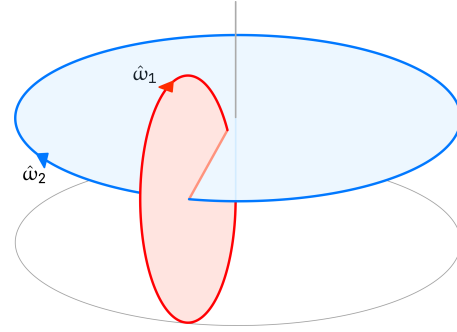
Figure 4:   Toroidal parameterisation of a complex lattice generated by a bi-periodic function
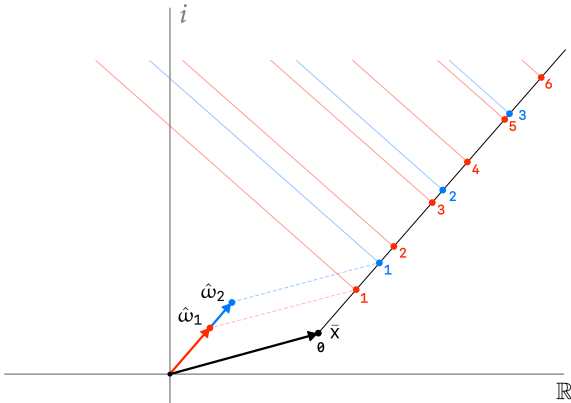

Figure 5:   Complex lattice generated by a mono-periodic function with linearly depen-dent complex periods
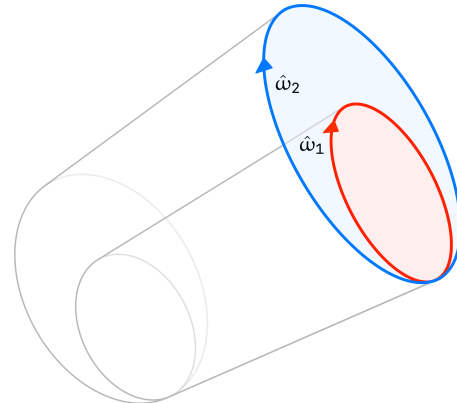

Figure 6:   Cylindrical parameterisation of a complex lattice generated by a mono-periodic function

We may now revisit Jacobi's trigonometric functions and understand how their topology allows for strict bi-periodicity. Like previously mentioned, Jacobi's functions are generalisations of generic sine and cosine functions which are traditionally defined on a circle. For instance, given any point (x, y) on a carterian surface, we may readily define the basic trigonometric functions to be

$$\sin\theta = \frac{x}{(x^2 + y^2)^{1/2}}, \cos\theta = \frac{y}{(x^2 + y^2)^{1/2}}.$$

This definition relies on the assumption that $(x^2 + y^2)^{1/2} = r$ is a constant equal to the radius of the circle on which the trigonometry is defined.

---

[2]Hyperplane refers to the (N + 1)-dimensional space that parameterises the periodic N-dimensional function

Jacobi's trigonometric functions extend these definitions to a generic path $r(\theta)$ other than a circle. This leads to a simple generation of sine and cosine to *sn* and *cn*, along with a third function *dn* that captures the variation in $r(\theta)$:

$$sn\,\mu \;=\; \frac{x}{r(\theta)}, \quad cn\,\mu \;=\; \frac{y}{r(\theta)}, \quad dn\,\mu \;=\; \frac{1}{r(\theta)}. \tag{10}$$

There are two main differences to note: first is the appearance of a third trigonometric function *dn* capturing the curvature of the path, and second is the change in argument from θ to μ. For the particular case of an elliptical path, μ can be evaluated by considering equation (8) for a generic final angle other than $\pi/2$ and through an equivalence to equation (5) such that

$$\mu \;\equiv\; C(e, \varphi) \;=\; \int_{0}^{\varphi} r(\theta, e)\,\delta\theta, \quad \text{where}$$

$$r \;\equiv\; r(\theta, e) \;=\; \frac{1}{(1 - e^2 \sin^2\theta)^{1/2}} \tag{11}$$

Jacobi noted that the trajectory (aka topology) of an ellipse is completely described by two concentric circles that inscribe (radius $\gamma = 1$) and circumscribe (radius α) the ellipse. HERE description of $sn^{-1}x$ function as shown in figures 7 and 8. In essense, Jacobi's functions are quite simply generation of trigonometry to
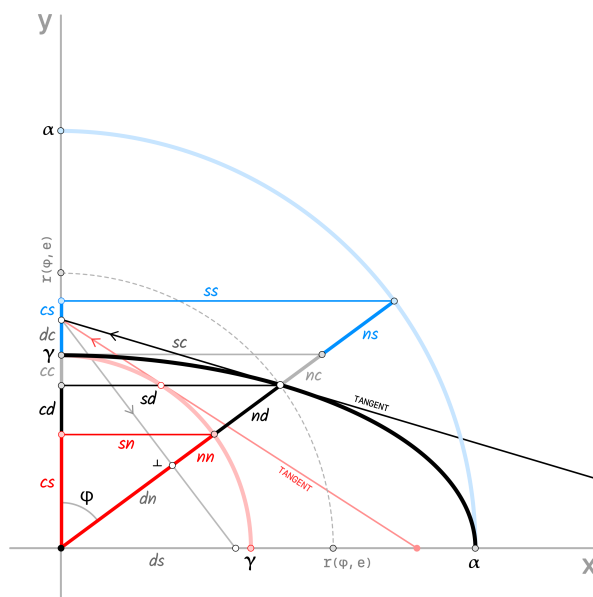

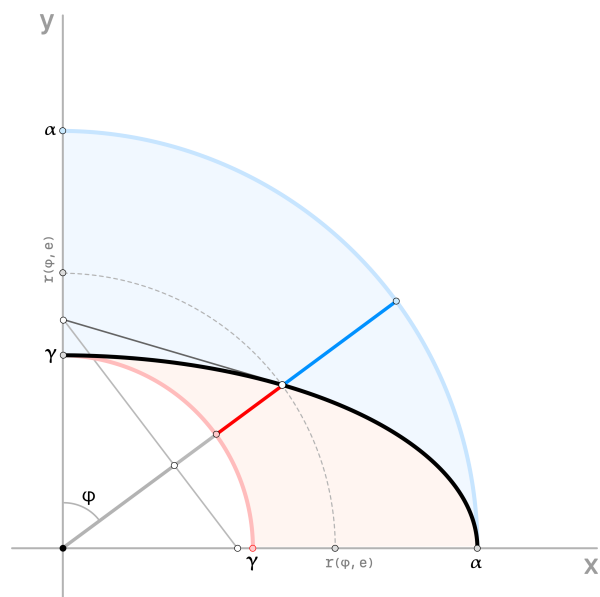Figure 7:   Jacobi's 16 trigonometric func-
tions

Figure 8:   Split topology for ellipses un-
derlying Jacobi's trigonometry

Fundamental theorems in vector algebra tell us that any two-dimensional vector space can have at most two independent basis vectors and any third vector can be written as a linear combination of those two bases. In the same way, any two-dimensional complex space is describable by at most two independent complex bases, in this case $\hat{\omega}_1$ and $\hat{\omega}_2$. There cannot be third period $\hat{\omega}_3$ since any such complex number will be reducible to $\hat{\omega}_3 = c_1\,\hat{\omega}_1 + c_2\,\hat{\omega}_2$ with constants $c_1, c_2 \in \mathbb{R}$. It is possible for two complex periods to be linearly dependent though, in which case the generated lattice is effectively pseudo-bi-periodic and its parameterisation yields two coaxial complex cylinders; a example of this is shown in figures 5 and 6. When there is only one complex period, the generated

lattice in that case is also strictly mono–periodic and forms a complex cylinder. This was the punchline of Jacobi's work: analytic functions can have either one complex period, or two (linearly) dependent complex bi–periods or two (linearly) independent complex bi–periods; among these three classes, only elliptic integrals appear to fall in the last category of independent bi–periods.

REFERENCES

[1] Intuitive Interpretation of Non–Interactive Zero–Knowledge Cryptography

METADATA