Around the Cubic Curve in Fifty Minutes[1]

William M. Faucette

University of West Georgia

When doing geometry—particularly when doing geometry over the field of complex numbers—it is most convenient to work in an ambient space which is compact. This requires that we work in projective spaces.

1 Preliminaries.

DEFINITION 1. *Let k be any field and let $\mathbb{A}^n = \mathbb{A}^n(k)$ be the set of all n-tuples in k. Projective n-space over k, written $\mathbb{P}^n(k)$. Projective n-space over k, written $\mathbb{P}^n(k)$, or simply $\mathbb{P}^n$, is defined to be the set of all lines through $(0, \ldots, 0)$ in $\mathbb{A}^{n+1}$. Any nonzero point $(x) = (x_0, \ldots, x_n) \in \mathbb{A}^{n+1} \setminus \{(0, \ldots, 0)\}$ determines a unique such line, namely $\{\lambda x_1, \ldots, \lambda x_n \mid \lambda \in k\}$. Two such points (x) and (y) determine the same line if $y_i = \lambda x_i$ for some nonzero $\lambda \in k$. Elements of $\mathbb{P}^n(k)$ will be called points. The point in $\mathbb{P}^n$ defined by the point $(x_0, \ldots, x_n) \in \mathbb{A}^{n+1} \setminus \{(0, \ldots, 0)\}$ will be denoted by $[x_0, \ldots, x_n]$, in which case the x_i's are called homogeneous coordinates.*

We remark that the affine plane sits naturally in the projective plane given by the map

$$\begin{aligned} \phi &: \mathbb{A}^2 \to \mathbb{P}^2 \\ \phi(X, Y) &= [X, Y, 1]. \end{aligned}$$

The remaining points in $\mathbb{P}^2$ form the *line at infinity*:

$$L_\infty = \{[X, Y, 0] \in \mathbb{P}^2\}$$

DEFINITION 2. *An affine plane curve is the zero set of a nonzero polynomial $F(X, Y)$ in $\mathbb{A}^2$. A projective plane curve is the zero set of a nonzero homogeneous polynomial $F(X, Y, Z)$ in $\mathbb{P}^2$. We say the curve has degree d if the defining polynomial has degree d. Curves of degree one, two, and three are called lines, conics, and cubics, respectively. For convenience, we will identify each curve—affine or projective— with its defining polynomial, so that we can speak of the projective curve F or $F(X, Y)$ or $F(X, Y, Z)$.*

DEFINITION 3. *Associated to each projective plane curve $F(X, Y, Z)$, there is a corresponding affine plane curve defined by the polynomial $F_*(X, Y) := F(X, Y, 1)$. This affine plane curve is simply the intersection of the projective plane curve with the embedded affine plane.*

DEFINITION 4. *If F defines a projective plane curve and $F = F_1 \cdot \ldots \cdot F_k$, where each F_i is irreducible, then the curves F_i are the (irreducible) components of the curve F. Such curves F_i are called irreducible.*

DEFINITION 5. *Let $\mathbb{A}^2(k) = \mathbb{A}^2$ be the affine plane over k. The local ring of $\mathbb{A}^2$ at P is*

$$\mathcal{O}_P(\mathbb{A}^2) = \left\{ \frac{F}{G} \in k(X, Y) \,:\, G(P) \neq 0 \right\}.$$

This ring is a local ring, i.e. it has a unique maximal ideal, in this case consisting of those rational functions vanishing at P:

$$m_P(\mathbb{A}^2) = \left\{ \frac{F}{G} \in k(X, Y) \,:\, F(P) = 0, G(P) \neq 0 \right\}.$$

DEFINITION 6. *Let F and G be affine plane curves and let $P \in F \cap G$. The intersection number $I(P, F \cap G)$ of F and G at P is*

$$I(P, F \cap G) = \dim_k \left(\mathcal{O}_P(\mathbb{A}^2)/(F, G) \right).$$

EXAMPLE 1. *Let $F(X, Y) = 3X^3 - 2X^2Y + XY - Y$ and $G(X, Y) = Y$. Then the intersection number at $P = (0, 0)$ is*

$$\left(\mathcal{O}_P(\mathbb{A}^2)/(F, G) \right) \cong k[X, Y]_{(x,y)}/(F, G) \cong k[X]_{(x)}/(X^3),$$

which has basis $\{1, X, X^2\}$ as a complex vector space, so $I(P, F \cap G) = 3$.

[1] Apologies to Jules Verne

Intersection numbers can also be defined for projective plane curves by choosing a line not containing the point of intersection, letting this line be the line at infinity, and looking at the corresponding affine plane curve.

THEOREM 1 (BEZOUT'S THEOREM). *Let F and G be projective plane curves of degrees m and n respectively. Assume F and G have no common component. Then*

$$\sum_P I(P, F \cap G) = mn.$$

Although we will not prove this theorem in it full generality, if G is a linear polynomial L, then $F \cap L$ is the intersection of the curve F and a line. Bézout's Theorem then immediately follows from the Fundamental Theorem of Algebra.

DEFINITION 7. *A point P on a affine curve F is called a simple point or smooth point if $F_X(P) \neq 0$ or $F_Y(P) \neq 0$. Otherwise, P is called a singular point or multiple point.*

These definitions extend to projective plane curves in the obvious way, with the condition for smoothness being $F_X(P) \neq 0$ or $F_Y(P) \neq 0$ or $F_Z(P) \neq 0$.

DEFINITION 8. *Let $P = (x_0, y_0)$ be a simple point on an affine plane curve $F(X, Y)$ The tangent line to F at P is the line*

$$F_X(x_0, y_0)(X - x_0) + F_Y(x_0, y_0)(Y - y_0) = 0.$$

DEFINITION 9. *A simple point P on a curve F is called a flex if $I(P, F \cap L) \geq 3$, where L is the tangent line to F at P.*

DEFINITION 10. *Let F be a projective plane curve of degree n. The Hessian of F is the polynomial*

$$H(X, Y, Z) = \det \begin{bmatrix} F_{XX} & F_{XY} & F_{XZ} \\ F_{YX} & F_{YY} & F_{YZ} \\ F_{ZX} & F_{ZY} & F_{ZZ} \end{bmatrix}$$

THEOREM 2. *(Assume char(k)$\neq 0$). A simple point P is a flex of a curve F if and only if the Hessian of F vanishes at P.*

PROOF: Let $P = (x_0, y_0)$ be a nonsingular point of a curve $F(X, Y)$. We will find the conditions that guarantee that P is an inflection point.

First, write the equation F in the form

$$F(X, Y) = a(X - x_0) + b(Y - y_0) + c(X - x_0)^2 + d(X - x_0)(Y - y_0) + e(Y - y_0)^2 + \text{ higher order terms}$$

Restricting F to the line L given by $X = x_0 + \lambda t$, $Y = y_0 + \mu t$, we have

$$F = (a\lambda + b\mu)t + (c\lambda^2 + d\lambda\mu + e\mu^2)t^2 + t^3\psi(t).$$

Therefore, the line L will have intersection multiplicity at least 3 if

$$\begin{aligned} a\lambda + b\mu &= 0 \\ c\lambda^2 + d\lambda\mu + e\mu^2 &= 0. \end{aligned}$$

The first of these equations says that

$$F_X(x_0, y_0)\lambda + F_Y(x_0, y_0)\mu = 0,$$

which is equivalent to L being the tangent line to F at P.

The second equation, taken together with the first, means the conic $cu^2 + duv + ev^2 = 0$ is reducible: It's divisible by $au + bv = 0$. In particular, the quadratic equation $g(u, v) := au + bv + cu^2 + duv + ev^2$ is reducible. The condition for this quadratic polynomial to be reducible is that

$$\det \begin{bmatrix} 2c & d & a \\ d & e & b \\ a & b & 0 \end{bmatrix} = \det \begin{bmatrix} g_{uu} & g_{uv} & g_u \\ g_{uv} & g_{vv} & g_v \\ g_u & g_v & 0 \end{bmatrix} = 0.$$

If we "homogenize" this last matrix, we have

$$\det \begin{bmatrix} F_{XX} & F_{XY} & F_{XZ} \\ F_{XY} & F_{YY} & F_{YZ} \\ F_{XZ} & F_{YZ} & F_{ZZ} \end{bmatrix} = 0,$$

which is the desired result. □

THEOREM 3. *Every cubic curve has nine flexes.*

PARTIAL PROOF: Let F be a nonsingular cubic curve and let H be the Hessian of F. Then F has degree 3 and H has degree $3(3-2)=3$, so by Bézout's Theorem, $\sum_P I(P, F\cap H)=9$. These are nine distinct points with intersection number 1 at each. □

2 Choosing the Right Coordinates.

Let C be a nonsingular cubic curve with defining polynomial F. Let $O\in C$ be a flex. By changing coordinates, we may take the tangent line at $O=[0,1,0]$ to be the line at infinity L_∞. Then

$$F(X,Y,Z)=ZY^2+bYZ^2+cXYZ+\text{ terms in } X,Z$$

By the map $Y\mapsto Y-\frac{b}{2}Z-\frac{c}{2}X$, which is an invertible linear transformation on $\mathbb{C}^3$ and therefore induces an automorphism of $\mathbb{P}^2$, we can put F in the form

$$ZY^2=\text{ cubic in } X,Z.$$

Setting $Z=1$, a nonsingular cubic curve has the affine equation

$$y^2=x^3+ax^2+bx+c.$$

If $k=\mathbb{C}$ we can do even better. First, set $Z=1$ so we're looking in the affine plane, $\mathbb{C}^2$. In the cubic polynomial on the right, there is a unique linear fractional transformation in the X coordinate taking the roots of the resulting cubic polynomial to zero, one, and some complex number $\lambda\neq 0,1$. Hence, we may assume the equation of F has the form

$$ZY^2=X(X-Z)(X-\lambda Z)$$

So, in the affine plane $\mathbb{C}^2$, a nonsingular cubic curve has the equation

$$y^2=x(x-1)(x-\lambda).$$

3 Families of Cubic Curves.

Let's consider the family of all cubic curves in the projective plane. This amounts to looking at the family of all homogeneous cubic polynomials in three variables. The typical such polynomial has the form

$$aX^3+bY^3+cZ^3+dX^2Y+eX^2Z+fYZ^2+gXY^2+hXZ^2+iY^2Z+jXYZ,$$

so we may view the family of such polynomials as the set of all 10-tuples of complex numbers, (a,b,c,d,e,f,g,h,i,j). However, polynomials which are constant multiples of one another represent the same projective curve, so we must identify all 10-tuples (a,b,c,d,e,f,g,h,i,j) which are nonzero multiples of one another, i.e. we must form a projective space.

As before, we will define a 10-tuple (a,b,c,d,e,f,g,h,i,j) to be equivalent to $(a',b',c',d',e',f',g',h',i',j')$ if $a'=\lambda a$, $b'=\lambda b$, $\ldots$, $j'=\lambda j$ for some non-zero complex constant λ. The resulting space is the set of all linear subspaces of dimension one in $\mathbb{C}^{10}$, a space which we will denote $\mathbb{P}^9$, nine dimensional projective space.

Suppose a cubic curve C with equation

$$F(X,Y,Z)=aX^3+bY^3+cZ^3+dX^2Y+eX^2Z+fYZ^2+gXY^2+hXZ^2+iY^2Z+jXYZ$$

passes through the point $[X_0,Y_0,Z_0]$ in the projective plane $\mathbb{P}^2$. Then

$$0=aX_0^3+bY_0^3+cZ_0^3+dX_0^2Y_0+eX_0^2Z_0+fY_0Z_0^2+gX_0Y_0^2+hX_0Z_0^2+iY_0^2Z_0+jX_0Y_0Z_0.$$

Viewing this as an equation in the variables a, b, c, d, e, f, g, h, i, and j with complex coefficients, we see that the family of cubic curves in the projective plane passing through a point is the zero locus of a linear homogeneous polynomial—it is a *hyperplane* in $\mathbb{P}^9$.

Now let's choose nine points in the projective plane $P_i=[X_i,Y_i,Z_i]$ and look at the family of cubic curves passing through all nine points. This family is then the intersection of nine hyperplanes in $\mathbb{P}^9$. As long as the conditions imposed on the projective space $\mathbb{P}^9$ are independent, that is, as long as the linear equations

$$aX_i^3+bY_i^3+cZ_i^3+dX_i^2Y_i+eX_i^2Z_i+fY_iZ_i^2+gX_iY_i^2+hX_iZ_i^2+iY_i^2Z_i+jX_iY_iZ_i=0,$$

are linearly independent, then the solution of this system of linear equations is a one-dimensional linear subspace of $\mathbb{C}^{10}$— a point in $\mathbb{P}^9$. This gives us the following proposition:

PROPOSITION 4. *Given nine general points in the projective plane $\mathbb{P}^2$, there is a unique cubic containing those points.*

In this same vein, suppose we have two projective plane curves C_1 and C_2, given by homogeneous polynomials F_1 and F_2, respectively, which pass through the nine points $P_1, \ldots, P_9$, which we assume are in general position. Suppose C is another cubic curve in the projective plane given by a homogeneous polynomial F and suppose C passes through $P_1, \ldots, P_8$.

We have seen that the family of homogeneous cubic polynomials in three variables is ten dimensional. Each point a cubic curve passes through defines one linear condition on this ten dimensional family. If the points are in general position, the family of cubic curves passing through eight points is then two dimensional. Since C_1, C_2, and C are all curves which pass through the eight points $P_1, \ldots, P_8$, the three polynomials F_1, F_2, and F must be linearly dependent. In particular, since F_1 and F_2 vanish at P_9 and F is linearly dependent on F_1 and F_2, we must likewise have that F vanishes at P_9. This gives the following proposition:

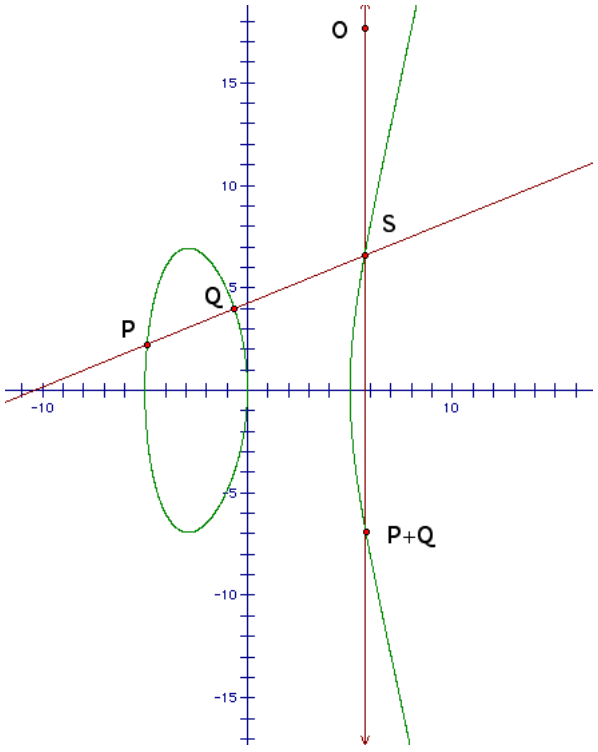
PROPOSITION 5. *Let C, C_1, and C_2 be three cubic curves. Suppose C goes through eight of the nine intersection points of C_1 and C_2. Then C goes through the ninth intersection point.*

4 Group Structure on a Nonsingular Cubic.

If we take two points, P and Q, on a nonsingular cubic C, construct the line $\overleftrightarrow{PQ}$. Since C has degree three, the line $\overleftrightarrow{PQ}$ meets C in one additional point S. This would appear to be something like a group, where you take two elements of some set, perform some operation, and get another element of that set. Unfortunately, the operation just described has no identity.

In order to remedy this problem, we modify the operation. Take two points, P and Q, on a nonsingular cubic C, construct the line $\overleftrightarrow{PQ}$ and let S be the third point of intersection of $\overleftrightarrow{PQ}$ and C. Now, construct the line $\overleftrightarrow{OS}$, where O is a fixed point on C. (Typically, O is chosen to be one of the nine flexes.) Next, we define the sum $P + Q$ be to the third point of intersection of the line $\overleftrightarrow{OS}$ with C. (See Figure 1.)

Figure 1.



THEOREM 6. *The operation just described gives an abelian group structure to the set of points on a nonsingular cubic curve.*

PROOF: Closure is clear since the sum of two points on the curve clearly produces another point on the curve. It is also clear that this operation is commutative since the line through two points is independent of the order of the points.

Next, we check that O is the additive identity. Let P be any point on the curve. Construct the line $\overleftrightarrow{OP}$ and let S be the third point of intersection of this line with C. Now, construct the line $\overleftrightarrow{OS}$. Noting that O, P, and S are collinear, it follows that the third point of intersection of the line $\overleftrightarrow{OS}$ with C is P. Hence, $P + O = P$, as desired.

Next, we check additive inverses. Let P be any point on the curve. First, we need a candidate for $-P$. Construct the tangent line $\overleftrightarrow{OO}$ to C and let T be the third point of intersection of this line with C. Construct the line $\overleftrightarrow{PT}$ and let U be the third point of intersection of this line with C. We claim that $-P = U$.

To see this, construct the line $\overleftrightarrow{PU}$. Noting that P, T, and U are collinear, we see that the third point of intersection of this line with C is T. Next, construct the line $\overleftrightarrow{TO}$. Since the line $\overleftrightarrow{TO}$ is the tangent line to C at O, the third point of intersection of this line with C is O. Hence, $P + U = O$, as desired.

The hardest part of this is to show that the operation is associative. Suppose P, Q, and R are points on C. Construct the line $L_1 = \overleftrightarrow{PQ}$ and let S' be the third point of intersection of this line with C. Construct the line $M_1 = \overleftrightarrow{S'O}$ and let S be the third point of intersection of this line with C. Next, construct the line $L_2 = \overleftrightarrow{SR}$ and let T' be the third point of intersection of this line with C. By definition, $(P + Q) + R$ is the third point of intersection of the line $\overleftrightarrow{T'O}$ and C.

On the other hand, construct the line $M_2 = \overleftrightarrow{QR}$ and let U' be the third point of intersection of this line with C. Now, construct the line $L_3 = \overleftrightarrow{U'O}$ and let U be the third point of intersection of this line with C. Next, construct the line $M_3 = \overleftrightarrow{UP}$ and let T'' be the third point of intersection of this line with C. Once again, by definition, $P + (Q + R)$ is the third point of intersection of the line $\overleftrightarrow{T''O}$ and C.

So, it suffices to show that $T' = T''$. See Figure 2.

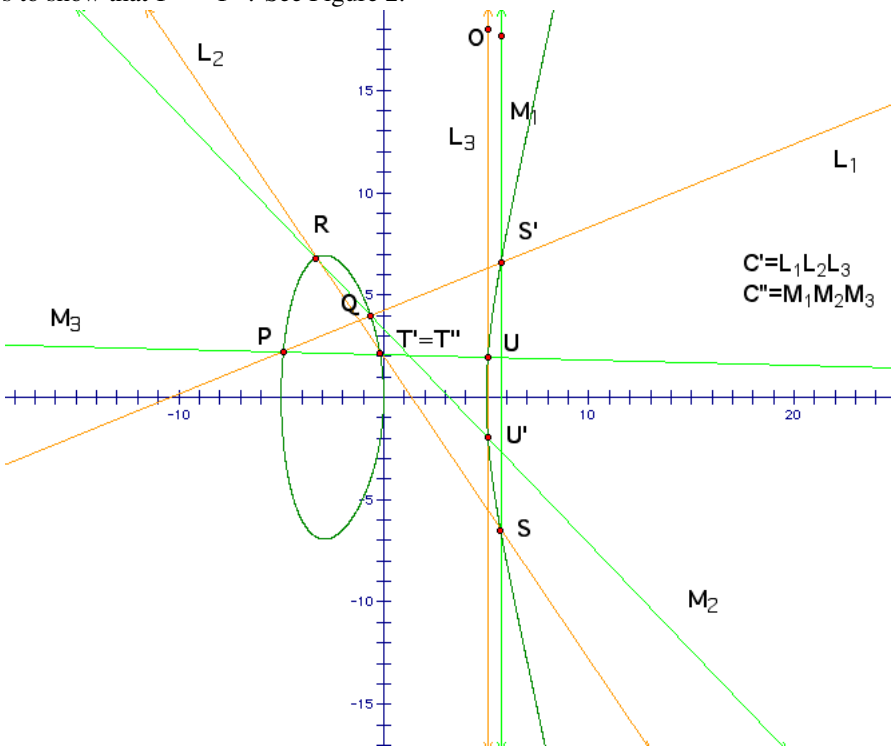


Figure 2.

Let C' be the reducible cubic $L_1L_2L_3$ and let C'' be the reducible cubic $M_1M_2M_3$. Then $C' \cap C = \{P, Q, R, O, S, S', T', U, U'\}$ and $C'' \cap C = \{P, Q, R, O, S, S', T'', U, U'\}$. Since C passes through eight points of the intersection of C' and C'', namely $\{P, Q, R, O, S, S', U, U'\}$, C must pass through the ninth point of intersection of C' and C''. It follows that $T' = T''$ and associativity follows immediately. $\square$

5 Explicit formulas.

We assume that the nonsingular cubic C has an equation of the form

$$y^2 = x^3 + ax^2 + bx + c$$

as shown earlier.

Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be points on C with $x_1 \neq x_2$. Let the third point of intersection of $\overleftrightarrow{PQ}$ be the point $P * Q = (x_3, y_3)$. By the choice of coordinates, $P + Q$ is the point $(x_3, -y_3)$.

Let $\lambda = \dfrac{y_2 - y_1}{x_2 - x_1}$ and $\nu = y_1 - \lambda x_1 = y_2 - \lambda x_2$. Then the line through P and Q has the equation $y = \lambda x + \nu$. To get the third point of intersection of the line $\overleftrightarrow{PQ}$ with C, we solve

$$(\lambda x + \nu)^2 = x^3 + ax^2 + bx + c.$$

By construction, two of the roots of this polynomial are x_1 and x_2, this polynomial factors as

$$(x - x_1)(x - x_2)(x - x_3),$$

so that the coefficient of x^2 is $-(x_1 + x_2 + x_3)$. This observation and a little algebra allows us to compute

$$x_3 = \lambda^2 - a - x_1 - x_2$$
$$y_3 = \lambda x_3 + \nu$$

Of course, these formulas make sense over any field. In particular, if we look over number fields, i.e. algebraic extensions of the field $\mathbb{Q}$ of rational numbers, you get the amazingly rich area in number theory called elliptic curves. A point on an elliptic curve is called a *rational point* if its coefficients are rational numbers. Undoubtedly, one of the central results here is

THEOREM 7 (MORDELL'S THEOREM). *Let C be a nonsingular rational cubic curve, then there is a finite set of rational points such that all other rational points can be obtained from the geometric construction described above. In particular, the group of rational points on C is a finitely generated abelian group.*

This result is now being used in cryptographical applications called *elliptic curve cryptography,* which is a collection of algorithms for using the group structure on an elliptic curve to encode sensitive data for electronic transmission.

6 Cubic Curves from a Different Perspective.

Earlier, we showed that a nonsingular cubic curve over the complex numbers can always be represented by an affine equation of the form

$$y^2 = x(x - 1)(x - \lambda).$$

If we consider the map $\pi_x : \mathbb{C}^2 \to \mathbb{C}$ given by projection onto the first coordinate: $\phi(x, y) = x$. If we look at the cubic curve from this perspective, the nonsingular cubic curve can be realized as a branched covering of the Riemann sphere, the branch points being the three roots of the cubic polynomial in x on the right side of the equation and the point at infinity.

This perspective gives rise to the rich area in complex variables called Riemann surface theory, the study of complex manifolds of complex dimension 1.

From the topological classification theorem of orientable surfaces (which every Riemann surface is), every Riemann surface is topologically the Riemann sphere with a finite number of handles attached. The number of handles, g, is called the *genus* of the Riemann surface. In the case of a nonsingular cubic curve, the genus is 1, so that the Riemann surface associated with a nonsingular cubic curve is $S^1 \times S^1$, at least topologically. Then, since it's clear that S^1 itself is a group, so $S^1 \times S^1$ is likewise a group. However, from this perspective, we don't get the beautiful geometric interpretation of the group structure.

7 Back to Algebraic Geometry.

On the other hand, every Riemann surface of genus 1 can be realized as a nonsingular cubic projective plane curve. To see this, we need a bit more terminology and one extremely major result.

DEFINITION 11. *Let C be a Riemann surface. A divisor on C is a formal sum*

$$D = \sum_{P \in C} n_P P, \, n_P \in \mathbb{Z} \text{ and } n_P = 0 \text{ for all but a finite number of } P.$$

DEFINITION 12. *Let f be a meromorphic function on a Riemann surface C. The divisor of f is*

$$(f) = \text{ zeroes of } f - \text{ poles of } f.$$

QUICK FACTS:

- Any meromorphic function on a Riemann surface has an equal number of zeroes and poles, counted with multiplicity. So, the degree of the divisor (f) is 0.
- It follows immediately that the only holomorphic functions on a Riemann surface are the constant functions.

DEFINITION 13. *Let ω be a holomorphic differential on a Riemann surface C. The divisor of ω is*

$$(\omega) = \text{ zeroes of } \omega - \text{ poles of } \omega.$$

QUICK FACTS:

- the degree of the divisor (ω) is $2g - 2$ on a Riemann surface of genus g. This divisor is called the *canonical divisor*. Hence, for a curve of genus $g = 1$, the degree of any canonical divisor is 0.
- More specifically, dz is a holomorphic differential on Riemann surface of genus $g = 1$. It has no zeroes and no poles, so $(dz) = (0)$.

DEFINITION 14. *Let D be a divisor on a Riemann surface C. The vector space $L(D)$ is*

$$L(D) = \{\text{meromorphic functions } f \mid (f) + D \geq 0\}.$$

To put this in simpler terms, if $D = \sum n_P P$ and all the $n_P \geq 0$, then a meromorphic function f is in $L(D)$ exactly when f has poles no worse than order n_P at P for each P in C. Let $\ell(D)$ denote the dimension of $L(D)$ over $\mathbb{C}$.

Unquestionably, one of the central results in the theory of Riemann surfaces is

THEOREM 8 (RIEMANN-ROCH THEOREM). *Let K be a canonical divisor on a Riemann surface C of genus g. Then for any divisor D on C,*

$$\ell(D) = \deg(D) - g + 1 + \ell(K - D).$$

Note that for a Riemann surface of genus $g = 1$, the Riemann-Roch theorem implies

$$\ell(nP) = n,$$

for any $P \in C$ and any natural number n. Here, $L(nP)$ is just the vector space of meromorphic functions with a pole at P no worse than order n.

Fix $P \in C$. Then

$$\ell(P) = 1$$

and since the constant functions are included in $L(P)$, we see that C has no meromorphic function f with a single, simple pole at P and holomorphic elsewhere.

Next,

$$\ell(2P) = 2,$$

so there exists a meromorphic function x on C with a pole of order 2 at P and holomorphic elsewhere.

Next,

$$\ell(3P) = 3,$$

so there exists a meromorphic function y on C with a pole of order 3 at P and holomorphic elsewhere. In fact, we can take y to be the derivative of x.[2]

Finally, since

$$\ell(6P) = 6,$$

and the functions x^3, x^2, x, y^2, y, xy, and 1 all lie in the vector space $L(6P)$, these seven functions must be linearly dependent, so there is a complex linear relation between these seven functions. This relationship (almost) demonstrates C as a (necessarily nonsingular) projective plane curve of degree 3.

This brings us back to where we started, which is a perfectly good place to end.

[2]We should mention in passing that the function we call x is the Weierstrass $\wp$-function $\wp(z)$ and the function we call y is the derivative $\wp'(z)$.