The *abc* conjecture, Part I

February 10, 2003

1. Rational functions on $\mathbf{P}^1$

Let $f : \mathbf{P}^1 \to \mathbf{P}^1$ be a rational function defined over the complex numbers, i.e.,

$$f(t) = \frac{a(t)}{b(t)},$$

where $a, b \in \mathbf{C}[t]$ are coprime polynomials with $b(t) \neq 0$.

The *degree* of f is $\max \deg(a), \deg(b)$.

This can be rewritten in homogeneous form as

$$f(x : y) = (a(x, y) : b(x, y)),$$

where $a, b \in \mathbf{C}[x, y]$ are homogeneous polynomials of the same degree d. These homogeneous and non-homogeneous forms are basically equivalent, so we can think of $\mathbf{C}(t)$ as the field of rational functions on $\mathbf{P}^1$.

If $P \in \mathbf{P}^1(\mathbf{C})$ and $f \in \mathbf{C}(z)$, we define $\mathrm{ord}_P(f)$ to be the order of zero or pole of f at P. If $P = z_0 \in \mathbf{C}$, then $\mathrm{ord}_{z_0}(f)$ is the unique integer m such that $f(z)/(z - z_0)^m$ is analytic and nonzero at z_0. For $P = \infty$, we define $\mathrm{ord}_\infty(f(z)) = \mathrm{ord}_0(f(1/z))$. It is easy to see that if $f = a/b$ in lowest terms, then $\mathrm{ord}_\infty(f) = \deg(b) - \deg(a)$.

For all points P, we have $\mathrm{ord}_P(fg) = \mathrm{ord}_P(f) + \mathrm{ord}_P(g)$ and $\mathrm{ord}_P(1/f) = -ord_P(f)$ for all rational functions f, g.

For $Q \in \mathbf{P}^1(\mathbf{C})$, we define the *valency* $e_Q(f)$ to be

$$e_Q(f) = \mathrm{ord}_Q(f - f(Q))$$

if $f(Q) \neq \infty$, and to be $\mathrm{ord}_0(1/f)$ if $f(Q) = \infty$.

A point Q of $\mathbf{P}^1(\mathbf{C})$ is called a *critical point* of f (or a *ramified point*) if $e_Q(f) > 1$. If f is nonconstant, then basic facts from complex analysis imply that Q is a critical point of f if and only if f fails to be injective in any neighborhood of Q.

A *critical value* of f is the image under f of a critical point.

Critical points and critical values are the analogues of ramified prime ideals (upstairs and downstairs) in the number field case.

The valency $e_{z_0}(f)$ can be interpreted when $z_0 \in \mathbf{C}$, $f(z_0) \in \mathbf{C}$ as the unique positive integer e such that

$$f(z) - f(z_0) = (z - z_0)^e g(z)$$

with $g(z)$ analytic and nonzero at z_0.

Exercise: Use the chain rule to show that for all $f, g \in \mathbf{C}(z)$ and all $Q \in \mathbf{P}^1(\mathbf{C})$, we have

$$e_{fg}(Q) = e_f(g(Q))e_g(Q).$$

Deduce in particular that if $\alpha, \beta \in \mathbf{C}(z)$ are Mobius transformations, then

$$e_{f\circ\alpha}(Q) = e_f(\alpha(Q))$$

and

$$e_{\beta\circ f}(Q) = e_f(Q).$$

The first fundamental fact about rational functions is that if $P \in \mathbf{P}^1(\mathbf{C})$, then $f(Q) = P$ always has exactly $\deg(f)$ solutions in $\mathbf{P}^1(\mathbf{C})$, counting multiplicities. More precisely:

Theorem 1. *Let $f \in \mathbf{C}(t)$ be a nonconstant rational function, and let $P \in \mathbf{P}^1(\mathbf{C})$. Then*

$$\sum_{Q\in\mathbf{P}^1(\mathbf{C})\,:\,f(Q)=P} e_Q(f) = \deg(f).$$

Proof. Exercise. □

This result should remind you of the formula $\sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}/\mathfrak{p}} f_{\mathfrak{q}/\mathfrak{p}} = [L : K]$ for number fields. Here the curve $\mathbf{P}^1$ plays the role of a number field, and points of $\mathbf{P}^1(C)$ play the role of maximal ideals of $\mathcal{O}_K$ (or more suggestively, of places of K).

In particular, we obtain the following corollaries:

Corollary 1. *If $f \in \mathbf{C}(t)$ is nonconstant, then the number of zeros of f equals the number of poles of f (counting multiplicities).*

Corollary 2. *If $f \in \mathbf{C}(t)$ is a nonconstant rational function and $P \in \mathbf{P}^1(\mathbf{C})$, then $\mathrm{ord}_Q(f) = 1$ for all $Q \in f^{-1}(P)$ iff $\#f^{-1}(P) = \deg(f)$.*

If $\mathrm{ord}_Q(f) = 1$ for all $Q \in f^{-1}(P)$, we say that f is *unramified above* P. Otherwise, we say that f is *ramified* above P.

It will be convenient to have a criterion for when a map f is ramified at a given point $Q \in \mathbf{P}^1(\mathbf{C})$. We do this under some additional hypotheses, and leave the general case for the reader to work out.

Lemma 1. *If $Q \neq \infty$ and $f(Q) \neq \infty$, then f is ramified at Q iff $f'(Q) = 0$.*

Proof. Recall that

$$f'(t) = \frac{b(t)a'(t) - a(t)b'(t)}{b(t)^2}.$$

Let $f(Q) = \gamma \in \mathbf{C}$. By definition, Q is ramified iff $e(Q) > 1$, which (since $Q \neq \infty$) occurs precisely when $a(t) - \gamma b(t)$ has Q as a multiple root, i.e., when $a'(Q) - \gamma b'(Q) = 0$. Substituting $\gamma = f(Q) = a(Q)/b(Q)$ and using the assumption that $b(Q) \neq 0$, we see that this happens iff $f'(Q) = 0$. □

In particular, we have:

Corollary 3. *A nonconstant function has only finitely many critical points.*

For each point $Q \in \mathbf{P}^1(\mathbf{C})$, define a normalized absolute value on $\mathbf{C}(t)$ by $\|f\|_Q = e^{-\operatorname{ord}_Q(f)}$. Then we have $\log \|f\|_Q = -\operatorname{ord}_Q(f)$, so that

$$\sum_{Q \in \mathbf{P}^1(\mathbf{C})} \log \|f\|_Q = 0,$$

which is the function field analogue of the product formula.

If we now consider the function

$$h(f) = \sum_{P \in \mathbf{P}^1(\mathbf{C})} \log \max\{\|f\|, 1\},$$

then we see that $h(f)$ simlply counts up the number of poles of f, with multiplicity. Therefore we have

$$h(f) = \deg(f).$$

So the "height" in our function field situation is just the degree of a rational function.

2. Riemann-Hurwitz and Mason's theorem

Another fundamental formula in function fields is the *Riemann-Hurwitz formula*:

Theorem 2. *If $f(z) \in \mathbf{C}(z)$ is a nonconstant rational function, then*

$$\sum_{Q \in \mathbf{P}^1(\mathbf{C})} (e_Q(f) - 1) = 2\deg(f) - 2.$$

As there are only finitely many critical points of f, the potentially infinite sum in the statement of the theorem is indeed finite.

Proof. Suppose first that $Q \neq \infty$, $f(Q) \neq \infty$. We want to express the quantity $e(Q) - 1$ in terms of the derivative $f'(z)$ of $f(z)$. For this, write $Q = z_0$, $e = e_Q(f)$, and then by definition,

$$f(z) - f(z_0) = (z - z_0)^e g(z)$$

where $g(z)$ is analytic and nonzero at z_0.

Then by the product rule,

$$f'(z) = e(z - z_0)^{e-1} g(z) + (z - z_0)^e g'(z),$$

so that $f'(z)/(z - z_0)^{e-1}$ is analytic and nonzero at z_0, and therefore $\mathrm{ord}_Q(f') = e_Q(f) - 1$ as desired.

If we write $f(z) = a(z)/b(z)$ with a, b coprime polynomials and recall that

$$f'(z) = \frac{h(z)}{b(z)^2},$$

where $h(z)$ is the polynomial $b(z)a'(z) - a(z)b'(z)$, then we can rephrase this as saying that if $Q \neq \infty$, $f(Q) \neq \infty$, then

$$\mathrm{ord}_Q(h) = e_Q(f) - 1. \tag{1}$$

If $Q \neq \infty$ but $f(Q) = \infty$, we claim that formula (1) holds as well. For this, we recall that in this case, $e = e_Q(f)$ is defined by saying that

$$f(z) = (z - z_0)^{-e} g(z)$$

with $g(z)$ analytic and nonzero at z_0. This is equivalent to having

$$b(z) = (z - z_0)^e \tilde{g}(z)$$

with $\tilde{g}(z)$ analytic and nonzero at z_0.

Therefore

$$f'(z) = -e(z - z_0)^{-e-1} g(z) + (z - z_0)^{-e} g'(z),$$

so that

$$h(z) = b(z)^2 f'(z) = -e(z - z_0)^{e-1} g(z)\tilde{g}(z)^2 + (z - z_0)^e g'(z)\tilde{g}(z)^2$$

and therefore $\mathrm{ord}_Q(h) = e_Q(f) - 1$ as claimed.

Since $\sum_{Q \in \mathbf{C}} \mathrm{ord}_Q(h) = \deg(h)$, it follows that we have

$$\sum_{Q \in \mathbf{P}^1(\mathbf{C})} (e_Q(f) - 1) = (e_\infty(f) - 1) + \deg(h). \tag{2}$$

It remains to calculate $e_\infty(f)$.

By definition, we have

$$e_\infty(f) = e_0(f(1/z)).$$

If we write $f(1/z) = \tilde{a}(z)/\tilde{b}(z)$, with $\tilde{a}(z), \tilde{b}(z)$ coprime polynomials, then by our previous calculations,

$$e_0(f(1/z)) - 1 = \mathrm{ord}_0(\tilde{h}(z)),$$

where

$$\tilde{h}(z) = \tilde{b}(z)\tilde{a}'(z) - \tilde{a}(z)\tilde{b}'(z).$$

On the other hand, if $d = \deg(f) = \max\{\deg(a), \deg(b)\}$, then it is easy to see that we can take $\tilde{a}(z) = z^d a(1/z)$ and $\tilde{b}(z) = z^d b(1/z)$. (The key point is that these polynomials must be coprime, which we leave as an exercise.)

Since

$$\tilde{a}'(z) = dz^{d-1}a(1/z) + z^d \frac{d}{dz} a(1/z) = dz^{d-1}a(1/z) + z^d(-1/z^2)a'(1/z)$$

and similarly for $\tilde{b}'(z)$, a simple calculation shows that

$$\tilde{h}(z) = -z^{2d-2}h(1/z),$$

so that

$$e_\infty(f) - 1 = \mathrm{ord}_0(\tilde{h}(z)) = (2d-2) - \deg(h).$$

The result now follows from (2). $\square$

Define the *radical* of a nonconstant rational function f to be

$$r(f) = \#f^{-1}(0) + \#f^{-1}(1) + \#f^{-1}(\infty).$$

We have as a consequence of Riemann-Hurwitz:

Theorem 3. *Let f be a nonconstant rational function. Then*

$$\deg(f) \le r(f) - 2.$$

An example which shows that this is sharp is the function $f(t) = t^n$, for which $\deg(f) = n$ and $r(f) = 1 + n + 1 = n + 2$.

Proof. Let $d = \deg(f)$. Noting that

$$\sum_{f(Q)=P} (e_Q(f) - 1) = d - \#f^{-1}(P),$$

we find that

$$\begin{aligned}
-2 &= -2d + \textstyle\sum_{Q\in\mathbf{P}^1(\mathbf{C})} (e_Q(f) - 1) \\
&\ge -2d + \textstyle\sum_{Q\in f^{-1}(0)} (e_Q(f) - 1) + \sum_{Q\in f^{-1}(1)} (e_Q(f) - 1) + \sum_{Q\in f^{-1}(\infty)} (e_Q(f) - 1) \\
&= -2d + (d - \#f^{-1}(0)) + (d - \#f^{-1}(1)) + (d - \#f^{-1}(\infty)) \\
&= \deg(f) - r(f).
\end{aligned}$$

$\square$

Now suppose $a(t), b(t), c(t)$ are nonzero polynomials. We define the height $h(a,b,c)$ of (a,b,c) to be the maximum of the degrees of a, b, c, and we define the radical $r(a,b,c)$ of (a,b,c) to be the number of *distinct roots* in $\mathbf{C}$ of the product abc.

We deduce the following theorem, which is the function field version of the *abc* conjecture.

Corollary 4 (Mason's theorem)**.** *If* $a(t), b(t), c(t) \in \mathbf{C}[t]$ *are relatively prime nonzero polynomials with* $a(t) + b(t) = c(t)$*, then*

$$h(a,b,c) \leq r(a,b,c) - 1.$$

A silly example which shows this is sharp is

$$t^n + 1 = (t^n + 1).$$

Proof. Let $f(t) = a(t)/c(t), g(t) = b(t)/c(t)$, so that $f(t) + g(t) = 1$. We have $\deg(f) = \max\{deg(a), \deg(c)\}$ by definition. Furthermore, the relation $b(t) = c(t) - a(t)$ precludes the possibility that $\deg(b) > \max\{\deg(a), \deg(c)\}$. Therefore $\deg(f) = \max\{deg(a), \deg(b), \deg(c)\} = h(a,b,c)$.

We now compare radicals. Since a, b, c are relatively prime in pairs, we have (for $t \in \mathbf{C}$) that $f(t) = 0$ iff $a(t) = 0$, $f(t) = 1$ iff $b(t) = 0$, and $f(t) = \infty$ iff $c(t) = 0$. Therefore $r(f) = r(a,b,c) + 1$ (resp. $r(f) = r(a,b,c)$) if $f(\infty) \in \{0, 1, \infty\}$ (resp. $f(\infty) \notin \{0, 1, \infty\}$). $\square$

As an application, the reader is invited to prove a version of Fermat's Last Theorem for polynomials in $\mathbf{C}[t]$.